AFRL-SR-AR-TR-09-0098

• • • • • • •• • • • • • • • • • • •• • •• • • • •	Form Approved OM8 No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
06-03-2009

2. REPORT TYPE
Final

3. OATES COVEREO *(From - To)*
1 December 2007 - 30 November 2008

4. TITLE AND SUBTITLE
Cognitive Coordination on the Network Centric Battlefield

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-0081 07-1-0081

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Cooke, Nancy J. & Myers, Christopher

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cognitive Engineering Research Institute
5810 South Sossaman Road, Suite. 106
Mesa, AZ 85212

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPDNSDRING/MDNITDRING AGENCY NAME(S) ANO AOORESS(ES)
Air Force Office of Scientific Research/PKR1
875 North Randolph Street, Room 3112
Arlington, VA 22203-1768

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR/PRK1

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Cognitive coordination is the timely and adaptive sharing of information. Command and control, particularly on the network-centric battle space, requires extensive coordination among a group of cognitive entities. Our goal was to better understand cognitive coordination and the impact of communication mode on team performance and coordination, and to develop a synthetic teammate. We conducted team experiments, and developed and tested a synthetic agent in the context of UAV (Unmanned Aerial Vehicle) ground control. An important objective of this was to closely match human behavior across several cognitive capacities, such as situation assessment, task behavior, and language comprehension and generation.
We also conducted an experiment in the UAV synthetic task environment to examine the coordination of a three-person team that interacted via voice- or text-based communications. Results demonstrated how text communications affect coordination relative to the voice communications. We plan to incorporate the synthetic teammate into human teams for validation, beginning summer 2009.

15. SUBJECT TERMS
Cognitive modeling, team cognition, team coordination, synthetic teammates

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	1B. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Nancy J. Cooke
UNCLAS	UNCLAS	UNCLAS	UL		19b. TELEPHONE NUMBER *(Include area code)* (480) 988-7306

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

20090325306

TABLE OF CONTENTS

TABLE OF FIGURES

COGNITIVE COORDINATION ON THE NETWORK CENTRIC BATTLEFIELD

Abstract

Cognitive coordination is the timely and adaptive sharing of information. Command and control, particularly on the network-centric battle space, requires extensive coordination among a group of cognitive entities (humans and agents). The goal of the proposed research was to better understand cognitive coordination and the impact of mode of communication, presence of synthetic teammates, and training regime on team performance and coordination.

We proposed a three-year effort to conduct two team experiments and to develop and test a synthetic agent in the context of UAV (Unmanned Aerial Vehicle) ground control. This work was motivated by theory and empirical research in team cognition and cognitive modeling, much of it attributed to research of the co-PIs on this project. An important objective of synthetic teammate development was to closely match human behavior across several cognitive capacities, such as situation assessment, task behavior, and language comprehension and generation. The initial application for the synthetic teammate research was the creation of an agent capable of functioning as the pilot of an Unmanned Aerial Vehicle (UAV) within a synthetic task environment (STE) which is described in the following section. At the same time, the work was motivated by a theoretical position that team member interaction that includes coordination is central to team cognition and a concomitant question concerning the role of the individual teammate in effective coordination. The ability of the synthetic teammate to participate in coordinated team interaction provides a systematic means to address this question.

In the two funded years of this project, we conducted an experiment in the UAV STE, developed infrastructure for the synthetic teammate modeling effort, and designed the synthetic teammate architecture.. Our experiment examined team coordination of a three-person UAV team that interacted via voice- or text-based communications. The text condition has demonstrated how this increasingly common form of interaction affects coordination relative to the voice condition and has provided data for developing language capabilities for the synthetic teammate. The goal of the project was to integrate task specific knowledge, a situation representation, and communication capabilities into a synthetic teammate capable of functioning as the Air Vehicle Operator (AVO), replacing the corresponding human teammate. However, funding for option two (year three) was not provided, thus the synthetic teammate was not integrated with humans performing the UAV task. Although funding from AFOSR has ceased, the synthetic teammate continues to be developed.

The Problem

The operational environment of today's U.S. Air Force is heavily dependent on command-and-control tasks that are increasingly cognitively-demanding, information-centric, and sensor-dependent in settings that are distributed, dynamic, uncertain, and fast-paced. The battlefield is not in any single geographic location, but is network-centric – distributed over a wide electronic web of sensors, cognitive agents, and effectors.

Generically, sensors push information to cognitive agents who filter, fuse, send it to other agents, and ultimately process it for action at the effector level (warfighters on the ground, weapons systems, other sensors). For the cognitive agent in this system, information overload is the rule. This network can be considered a cognitive system with inputs from the environment, processing, and outputs back to the environment. This military scenario has parallels in many civilian tasks including the response to hurricane Katrina, emergency operations centers, telemedicine, and air traffic control.

How can we assess performance of this cognitive system? Is the cognition of this system reflected in the collection of cognition of its individual cognitive agents? What factors influence decision-making, problem solving, and situation awareness at the level of the cognitive system? How can cognition be measured at the system level? How can we design for and train this cognitive system? How can we model cognition at this level? Our research program in the CERTT (Cognitive Engineering Research on Team Tasks) Lab is focused on these and other questions pertaining to team or collaborative cognition.

In particular, we are now focusing on coordination in cognitive systems. We define cognitive coordination as the timely and adaptive push and pull of information across the system. Our current focus on coordination is based on eight years of empirical data collected in our UAV command-and-control test bed (i.e., CERTT UAV-STE) that suggests that 1) teams learn; their performance improves even after reaching criterion on individual tasks; 2) shared knowledge of the task and team tend to improve initially, but cannot account for the acquisition of team skill; 3) team process, situation assessment, and communication do change with improvements in team performance; and 4) new team members or long retention temporarily hurt performance, but improve process and situation awareness over the long run. Team process, team situation awareness (assessed by our CAST (Coordinated Awareness of Situations by Teams) measure) and team communication are all highly relevant to coordination. In addition to leading us to a focus on team coordination, these data have caused us to rethink our theoretical approach to shared cognition. We now view team cognition more ecologically, as an emerging property of collaboration that is not reducible to the cognition of the individuals involved.

Across the funded two-year effort we more deeply examined team coordination in our UAV testbed. We conducted an experiment and began developing a computational model. Our empirical findings have driven the modeling of an ACT-R based synthetic teammate (i.e., the AVO). The experiment has been used to examine the impact of text-based communications on team cognition, and drive development of the synthetic teammate.

The synthetic teammate has the potential for a wide space of uses including its application as a training partner for teams and its pragmatic use as a reliable and easy to control teammate for empirical team studies. More specifically, in regard to training, our previous findings have led us to test two varieties of coordination training. One variety relies on prescribed coordination patterns that are imposed on the team through training and AARs (after Action Reviews). Our previous data indicate that prescribed coordination training should result in rapid acquisition of the prescribed coordination skill, but minimal flexibility. Alternatively, perturbed coordination training provides a rich array of situations to the team in which coordination patterns of the team must

change in concert with the situation. Our data suggest that teams develop flexibility in coordination with perturbed training. The synthetic agent will have the capacity to provide increased control over coordination training because it can be used to push and pull information in a prescribed manner. For example, our current teams of three individuals typically display enormous variance in their behavior, including coordination behavior. The addition of the Synthetic AVO will reduce the degrees of freedom in team behavior by constraining the behavior of the other team members. Therefore, the development of the Synthetic Teammate will not only break new ground in the computational modeling of teammates, but will also facilitate coordination training and experimental control.

The following objectives were identified as integral to the development of a synthetic agent acting as a teammate in the CERTT UAV-STE. Numbers in parentheses designate year in proposed 3-year effort. Because funding was discontinued in our third year, we focus this report on tasks identified for Year 1 or Year 2.consequently tasks identified below that contain (3) were not started or completed.

OBJECTIVE 1.0: Conduct Empirical Study of Cognitive Coordination to Guide Development of Synthetic Teammate

- Task 1.1 Modify synthetic test bed to accommodate chat-only communications (1)
- Task 1.2 Design Experiment 1 (chat vs. voice communications) (1)
- Task 1.3 Collect Experiment 1 data (1)
- Task 1.4 Analyze and report Experiment 1 (2)

OBJECTIVE 2.0: Develop Synthetic Teammate

- Task 2.1 Conduct task analysis of AVO performing reconnaissance task (1)
- Task 2.2 Develop plan for staging Synthetic AVO development for mitigation of risk (1)
- Task 2.3 Develop an interface between the CERTT simulation environment and ACT-R/Lisp (2)
- Task 2.3.1 Visual input to Synthetic AVO (2)
- Task 2.3.2 Data interface to support reimplementation of AVO GUI in ACT-R/Lisp environment (2)
- Task 2.3.3 Motor output from Synthetic AVO (2)
- Task 2.4 Develop Cognitive Model (reconnaissance task, cognitive control, reading, typing, comprehension of situation, cooperative dialog, representing other minds) (2-3)

OBJECTIVE 3.0: Conduct an Empirical Study to Validate Synthetic Teammate and Test Coordination Training

- Task 3.1 Incorporate Synthetic Teammate in synthetic test bed (2)
- Task 3.2 Design Experiment 2 (validation experiment, precise form depends on results of Experiment 1 and resulting features of synthetic teammate) (3)
- Task 3.3 Collect Experiment 2 data (3)
- Task 3.4 Analyze and report Experiment 2 (3)

Background

In the following sections we provide requisite background that motivated the proposed research.

Team Cognition

One of the most common frameworks for conceptualizing team cognition puts shared mental models at the forefront of an I-P-O (input-process-output) framework. Applying the I-P-O framework to cognition at the team level is analogous to the information processing view of cognition at the individual level insofar that knowledge structure is distributed over team members, instead of over long term memory, and is operated on by team process behaviors, instead of memory processes. A generic I-P-O framework is presented in **Error! Reference source not found.**.

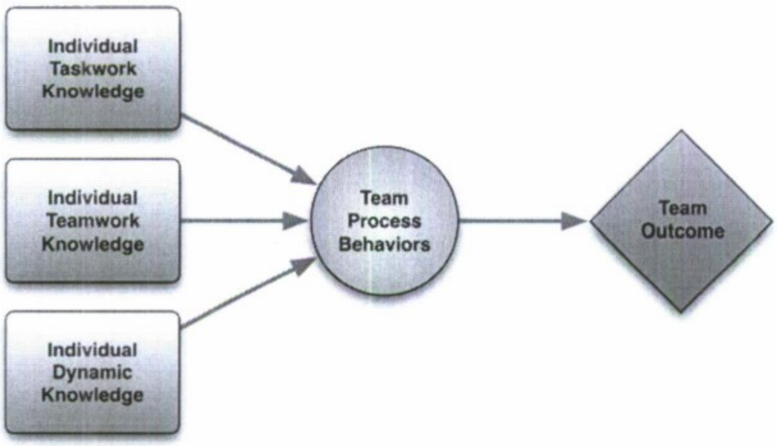

Figure 1. A generic Input-Process-Output (I-P-O) framework.

Interestingly, within this framework some have conceptualized team cognition as an outcome (e.g., Mathieu, et al., 2000). Others have considered collective cognition as an input in the I-P-O framework (e.g., Mohammed & Dumville, 2001) and others have viewed team cognition in terms of process behaviors such as planning and decision-making (e.g., Brannick, et al., 1995). So team cognition can and has been associated with all parts of the I-P-O framework, however, there has been increasing focus on the "I" part in which team *cognition is thought of as the collection of individual team member knowledge involving the task and team* (Figure 2, Panel A.)

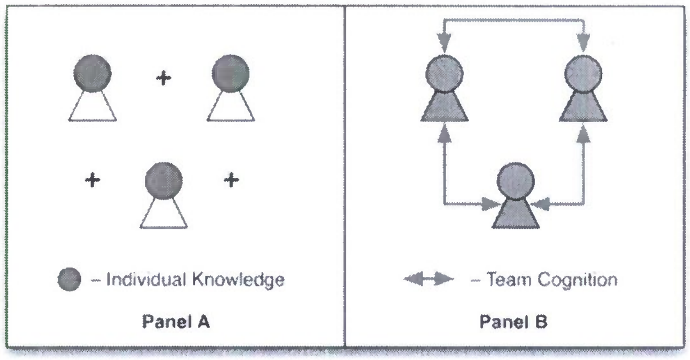

Figure 2. Team cognition as viewed from the collective (Panel A) and interaction (Panel B) perspectives.

Views of *shared mental models* and *team situation awareness* as common understanding, vision or knowledge across team members and the concomitant emphasis on knowledge in cognitive theories of individual expertise (Cooke, 1994) turned the spotlight toward the input side of the I-P-O framework. The focus was on the knowledge or mental models and not the sharing processes. For example, these sharing processes have been tied to knowledge (e.g., Entin & Serfaty, 1999). Thus the information processing perspective is knowledge-centric, rather than behavior-centric (e.g., Mohammed & Dumville, 2001). At the same time, with this emphasis also came a shift from decentralized notions of adaptive team coordination (cf. Tushman, 1979) to a more knowledge-homogeneous, static view.

We (i.e., the CERTT Lab team) have conceptualized team cognition differently. We take an alternative perspective to the I-P-O framework that is partially motivated by some limitations of the IP perspective (i.e., applicability to heterogeneous teams, knowledge vs. process focus) and partially motivated by some alternative views of scientific psychology (i.e., distributed cognition, Hutchins, 1991; ecological psychology, Reed, 1996; dynamical systems theory, Alligood, Sauer, & York, 1996; and Soviet-era activity theory, Leontev, 1990). This ecological view considers team cognition as emergent, rather than a linear aggregate, and is thus focused on the dynamic interactions among team members, rather than the static structure of team member knowledge. It is accordingly, a perspective on team cognition that supports interaction rather than aggregate measurement. As represented in Figure 2, Panel B, *team cognition is not equivalent to the linear aggregate of individual team member cognition, but instead emerges from the dynamic interactions among teammates.*

This perspective advocates thinking about and measuring teams at the team level of analysis rather than measurement of individuals (and aggregation) and is inspired by Gestalt psychology (Cooke, et al., 2000; see also "collective cognition," Gibson, 2001). Simple aggregation rules (e.g., summing) are inappropriate for heterogeneous teams for which there is a heterogeneous distribution of knowledge and abilities across team members (Cooke & Gorman, in press; Gorman, Cooke, & Kiekel, 2004). In an aggregate the parts are independent of their relations to each other while in a whole, relations help determine the nature of the parts. For interaction team cognition the relations among the parts are of inherent interest, in addition to the static distribution of knowledge among the

parts themselves. The ecological view is concerned with the team processing mechanisms by which the *whole* team is structured, beyond the sum of the parts. This emphasis on team member interactions beyond a collection of team knowledge stores is also shared with much of the small group work on decision-making (Festinger, 1954; Steiner, 1972), social decision schemes (Davis, 1973; Hinsz, 1995; 1999), and even transactive memory with its emphasis on transaction or communication (Hollingshead & Brandon, 2003).

Borrowing concepts from ecological psychology, teams can be viewed as a set of distributed perception-action systems that can become coordinated to the relatively global stimulus information specifying a team-level event. By analogy, when we encounter fire we see flames, we smell smoke, we feel the heat, we hear the crackle, etc.; our perceptual systems are coordinated to the same stimulus information specific to fire. Similarly, when an event occurs in the team environment, each team member is heterogeneously attuned to different aspects of the event. These "perception-action" systems are all attuned to the same event, they just extract information about it in different ways, in such a manner that these systems need to be coordinated. Our preferred perspective thus emphasizes team coordination (i.e., a team process) in response to events in the team environment. In this manner, team cognition is characterized as a single organism, ebbing and flowing and adapting itself to novel environmental constraints through the coordination of a team's perceptual systems. This process of adaptation is also consistent with activity theory (Leontev, 1990) or how a team internalizes new information in terms of information distribution across team members (cf. Artman, 2000).

In contrast to I-P-O-oriented theories of team cognition in which regression is used to predict team outcome at a single point in time, the ecological perspective considers the dynamic evolution of the "team as a system" using dynamical systems theory (Alligood, et al., 1996; e.g., Losada & Heaphy, 2004). For example, the concepts of circular causality, self-organization, bifurcation theory, and entrainment derived from dynamical systems theory are consistent with these views (Cooke and Gorman, in press; Gorman et al., submitted). This goes back to the early conceptualizations of team cognition and the realization that coordination is dynamic, not static, and has to *continually evolve* in order to handle the flux of information in highly complex team environments.

In our most recently funded work and in the proposed work we took the ecological perspective in terms of our approach to understanding and modeling team coordination and its development. However, we did not rule out the benefits of taking a more individualistic perspective on team cognition. Indeed our incorporation of ACT-R as an AVO agent reflected this perspective and we planned to show how these two views can co-exist as team cognition and coordination is examined at different levels of analysis.

Coordination

Coordination refers to the dynamic organization of diverse events and/or task elements in order to accomplish a task. Coordination further refers to patterned sequences of events. For example, random sequences of events are not likely to be coordinated. For the most part coordination has been studied in two distinct ways. First is what we call "blueprint" coordination. According to blueprint theory the study of coordination consists of

"characterizing different dependencies [between activities] and identifying the coordination processes that can be used to manage them" (Malone & Crowston, 1994, p. 91). Ideally in blueprint theory a "handbook" (Malone & Crowston, 1994, p. 92) of coordination processes would facilitate the understanding of coordinative phenomena in general. Conversely what we refer to as "emergent" coordination emphasizes that coordination is a naturally occurring phenomenon. That is, coordination can arise in systems without the aid of a blueprint or executive controller whenever they process information (or energy) via functional interactions between system elements (e.g., Kelso, 1995). According to blueprint theory, coordination processes govern element interactions and behavior at the individual level. In the emergent behavior theory the coordination level emerges from interactions between elements at the individual level. The emergent coordination level in turn influences behavior at the individual level by constraining interactions. In Figure 3, the left panel portrays blueprint coordination and the right panel portrays emergent behavior coordination.

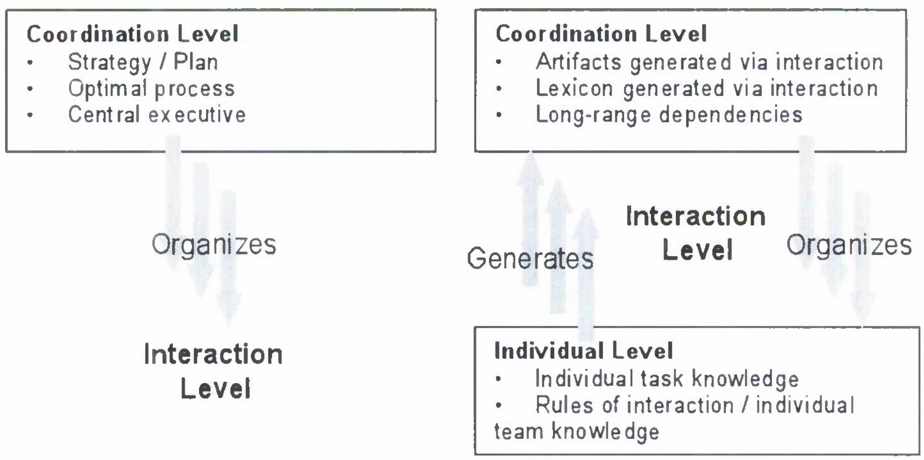

Figure 3. The left panel characterizes "blueprint" coordination; the right panel characterizes "emergent" coordination.

Our research on team coordination combines aspects of both blueprint coordination (LOM: Local Optimal Model) and emergent coordination (LOM dynamics). Specifically a LOM of coordination at salient target events was generated over the individual and interaction levels, and emergent patterns were identified in long trial sequences of LOM variability by human UAV teams. The fact that teams varied in their employment of the LOM from target-to-target in a patterned way suggests that for the UAV task the coordinative level is best described using an emergent behavior theory of team coordination. However the application of blueprint theory, in developing the LOM, is indispensable because it allows us to sample functional variation in target-level coordination processes.

Computational Cognitive Modeling

Research in computational cognitive modeling within cognitive architectures has reached the stage at which researchers are beginning to investigate the integration of models of

different cognitive components into more complex, integrated cognitive systems (Gray (ed.) in press; Cassimatis 2005; Cassimatis et al. 2004; Lee & Anderson 2001; Salvucci et al. 2001; Schoelles & Gray 2000; Scolaro & Santarelli 2002). Among the cognitive architectures being used to build complex cognitive systems are ACT-R (Anderson et al. 2004; Anderson & Lebiere 1998), SOAR (Newell 1990; Laird et al. 1987), EPIC (Kieras & Meyer 1997), Polyscheme (Cassimatis 2002), COGENT (Cooper 2002), ICARUS (Langley in press), CogNet (Zachary & Ross 1991) and CLARION (Sun 2005). Of these, ACT-R and SOAR have the largest user base, although the CogNet cognitive architecture is most closely associated with the development of synthetic teammates (Chapman et al. 2004).

The commitment to the use of a cognitive architecture to build complex cognitive systems is fueled by extensive empirical research, motivated by the goal of closely modeling human cognitive behavior, and is consistent with basic principles of cognitive science. This approach can be contrasted with Artificial Intelligence approaches aimed at the creation of intelligent systems without concern for cognitive plausibility. Both approaches may lead to development of intelligent systems capable of interacting with humans. From the perspective of AI and computer science, the constraints imposed by cognitive architectures may appear overly restrictive and unnecessary, reducing the chances for success. But computational cognitive modelers embrace cognitive constraints willingly and go even further in attempting to validate their cognitive models against fine grained human data (Gluck & Pew (eds.) 2005).

Why isn't the modeling of input/output behavior—however this is accomplished computationally—considered adequate within the cognitive modeling community? For one, it is because the current state of knowledge about human cognition, and the availability of cognitive architectures, permits us to model human behavior at a finer level of granularity. For another, it is because AI and computer science programs have proved inadequate to model human input/output behavior on complex tasks and in complex domains. The success of chess playing programs (or expert systems) is not an exception. No competing chess master would take the chess playing program for a human, although the program might ultimately win. If the goal is to develop programs capable of complex behavior, the use of cognitively implausible AI techniques and the adoption of a black box approach to human cognition may be acceptable or even preferable. If the goal is to develop cognitive models of complex human behaviors—as in the case of a synthetic teammate intended as a substitute for a human teammate in a training simulation environment—we need to look inside the black box of human cognition (cf. Ball, 2006). As Langley (in press) notes, this was originally an important goal of both AI and cognitive science.

Despite the availability of cognitive architectures for complex cognitive systems development, the research challenges are formidable. In complex cognitive systems, individual cognitive components often interact with each other in complex ways that are difficult to predict and model. Strong modularity (Fodor, 1983) is not a common feature of higher level cognitive processes and even weak modularity is difficult to reconcile with fMRI and other brain scanning evidence which suggests widespread brain activation during the performance of most any cognitive task. It is not that specific brain circuits are

not activated during specific cognitive tasks, rather it is that such brain circuits are not *exclusively* activated. Recent evidence that the same brain regions in the visual cortex are activated by the processing of spatial expressions as well as when performing spatial tasks (e.g. mental rotation), further weakens the modularity hypothesis (cf. Carpenter et al. 1999). Besides the weakening of the modularity hypothesis, the massive parallelism of the brain and the hierarchical cortical column structure on which this parallelism is manifested, means that many cognitive processes run in parallel with and on top of each other, making it difficult to tease them apart in terms of their behavioral manifestation. This is especially true of higher level cognitive processes which are often disguised by the lower level perceptual and motor processes with which they run in parallel, and which are closer to the input and output behaviors which can be experimentally measured (e.g. reaction time). Finally, the reality is that higher level cognitive constructs like attention (Pashler 1998) and (short-term) working memory (Ericsson & Kintsch, 1995) have proved difficult to localize in the brain. If the research of dynamic systems theory holds sway (Busemeyer 2002; Juarrero 1999; Holland 1998), these cognitive constructs may turn out to be emergent properties of neural networks (viewed as complex, dynamic systems) which cannot in principle be mapped to specific neural elements (although perhaps they can still be mapped to larger brain regions based on brain lesion, brain scanning and other evidence).

Even if the above theoretical issues can be overcome, the practical realities of building complex computational systems—whether cognitively motivated, or not—must also be overcome. It is well understood within AI and computer science, that solving most hard problems means avoiding the combinatorial explosion that results from the attempted use of algorithmic search techniques over large solution spaces. Many hard problems have simple algorithmic solutions that would take longer than the age of the universe to execute. By contrast, the human brain has evolved processes to solve many hard problems, returning reasonable solutions in real-time. Computational cognitive modelers interested in building complex cognitive systems will need to address issues of combinatorial complexity in ways that are compatible with what we know about how the human brain accomplishes this feat—as reflected in the modeling constraints imposed by cognitive architectures.

Another computational technique for solving hard problems is to break the problem down into simpler problems that can be solved in isolation and integrated to provide an overall solution. The modularity hypothesis was especially attractive because it provided theoretical support for adopting this approach in the development of cognitive systems. The modularity hypothesis also provided support for the purported existence of an autonomous syntax component (Chomsky 1965). Unfortunately, the empirical evidence does not support either strong modularity or the existence of an autonomous syntax component (Marslen-Wilson & Tyler 1987; Karmiloff-Smith 1992). Recent empirical investigations within the Visual World Paradigm (Trueswell & Tanenhaus 2004; Henderson & Ferreira 2004; Tanenhaus et al. 2000; Magnuson et al. 1999) demonstrate an extremely close relationship between the word-by-word processing of linguistic input and eye movements to a visual scene corresponding to the linguistic input. Humans fixate objects in the visual scene as soon as the linguistic input provides sufficient information

to discriminate among the objects. Further, from a computational perspective, modular approaches only work when the modules can be sufficiently isolated from each other.

With respect to language, the idea that syntactic analysis can first be performed within an autonomous syntax module impervious to higher level cognition with only the output of the syntactic analysis being made available for semantic analysis and higher level cognitive processes has turned out to be computationally intractable. Pervasive lexical and structural ambiguity—from a purely syntactic perspective—make it virtually impossible to construct a valid syntactic representation in the absence of higher-level semantic information. Unless the construction of a syntactic representation is constrained by semantic information, the likelihood of arriving at a correct syntactic representation in isolation is extremely low in any non-trivial system. Despite this rampant structural ambiguity, there is little evidence that humans are consciously aware of it. The best available explanation for this lack of awareness is that humans integrate syntactic and semantic information in a way that arrives at a reasonable interpretation of the input without explicit consideration of all the possible structural alternatives. Implicit probabilistic mechanisms executing in parallel perform much of the computation that ultimately leads to explicit awareness of the meaning of the input. There is no sharp divide between the implicit computations and the explicit representations that arise from them. Complex cognitive models will only be able to make limited use of modularization—just to the extent that such modularization is empirically justified. Instead, the key concern is in figuring out how to integrate lower level stochastic or rational mechanisms with higher level symbolic processes into coherent cognitive systems (Sun 2001; Wermter & Sun 2000).

Although the theoretical and computational challenges for developing complex cognitive systems are substantial, members of the Performance and Learning Models (PALM) team—two of whom were direct contributors to the proposed research—have extensive experience using the ACT-R cognitive architecture to develop computational cognitive models.

Developing Synthetic Teammates within Cognitive Architectures

The goal of the proposed research was to integrate task specific knowledge, situational awareness, and communication capabilities into a synthetic teammate capable of functioning as the AVO teammate in the Cognitive Engineering Research on Team Training (CERTT) testbed, replacing the human AVO teammate. The basic research associated with the computational modeling part of the proposed effort is on the exploration of theoretical and computational issues involved in the creation of a complex cognitive system within the context of the CERTT testbed and development of a synthetic AVO teammate.

The development of a synthetic entity capable of functioning as a teammate in a complex training simulation environment requires the integration of multiple cognitive components, each of which is a major topic of research in its own right. The synthetic agent must be capable of performing the task at hand and must be capable of communication and coordination with other teammates. To perform the task, the synthetic agent must interact with a GUI to encode relevant information, compare the

encoded information to the desired settings, make adjustments if necessary and be capable of responding to unexpected events. To communicate with human teammates using text-based communications, the synthetic agent must be able to comprehend incoming communications and type appropriate responses. From visual and text-based inputs, the synthetic agent must be capable of building a situation representation (Kintsch, 1998, Zwann & Radvansky, 1999) that takes into account the different perspectives of the other teammates. The situation representation is the basis for grounding linguistic representations, providing context for comprehension. The situation representation replaces abstract concepts as the basis for providing meaning to linguistics expressions. With purely abstract concepts with no perceptual basis banished, the brain can be viewed as a highly evolved perceptual (motor) organ (Barsalou, 1999; Prinz, 2002). Linguistic inputs lead to generation of perceptually based linguistic representations whose meanings are grounded in perceptually based representations of the objects and situations to which the linguistic expressions refer. The reason spatial expressions activate areas of the visual cortex involved in spatial processing, is because the meaning of spatial expressions resides in these same spatial processing regions. Just as mental imagery can activate spatial regions in the absence of any external input (Kosslyn 1994), so spatial expressions can activate these same spatial regions (Carpenter et al. 1999).

Overview of Research Effort

The ultimate goal of the research reported here is the development of a synthetic teammate capable of functioning as the AVO in the CERTT testbed. The synthetic teammate will make research on the use and effects of synthetic teammates to train team coordination possible. We are committed to the use of the ACT-R cognitive architecture to support this development. We are also committed to validating the synthetic teammate against fine-grained human performance data. To achieve theses goals we established the three following objectives:

1. Conduct an empirical study of cognitive coordination to guide the development of the synthetic teammate.
2. Develop synthetic teammate.
3. Conduct an empirical study to validate synthetic teammate and test coordination training.

Each of the above objectives was identified with a specific year of funding. For instance, the first objective was planned for the first year; the second objective was planned for option year two, etc. Because funding was discontinued for option year three, we have yet to incorporate the synthetic teammate into human teams to validate the synthetic teammate; however, development is continuing through other funding agencies and there are plans to incorporate the synthetic teammate into human teams by Winter 2009. The remaining sections of this report cover progress made in years one and two.

Synthetic Task Environment

The task environment used for developing the synthetic teammate is the Cognitive Engineering Research on Team Tasks (CERTT) UAV-STE (Unmanned Aerial Vehicle-

Synthetic Task Environment) (Cooke & Shope, 2005). The CERTT UAV-STE simulates teamwork aspects of UAV operations rather than equipment aspects (e.g., buttons and dials). The UAV-STE involves three interdependent team members, each with a different role. The team members are the Data Exploitation Mission Planning and Communications operator (DEMPC, the planning officer) who is responsible for a dynamic flight plan, including speed and altitude restrictions, an Aerial Vehicle Operator (AVO, the pilot) who controls flight settings and systems, and a Payload Operator (PLO, the sensor operator) who monitors sensor equipment and takes photographs.

The team members' common goal is to photograph ground targets and this requires interaction between all team members. Interaction occurs through a voice- or text-based communications system. A single UAV-STE mission consists of 11-12 ground targets and lasts a maximum of 40 minutes. However, a mission can end once the team photographs all possible targets.

The task requires a high degree of coordination due to time pressures and mutual constraints among the team member roles. To perform well within the UAV-STE, team members must understand their own tasks, and, more importantly, coordinate with each other to complete their common goal. The UAV-STE therefore provides an ideal task environment for developing a synthetic teammate.

Experiment

The purpose of the experiment is to determine how different communication modes affect team behaviors and processes within the UAV-STE. The two modes of communication that were used were text- and voice-based communications. Up until this point, voice over headsets using a push-to-talk intercom systems was the primary mode of communication in the UAV-STE. In this project we switched to text-based (i.e., "chat") communications. Chat communications relieve the synthetic teammate of speech recognition requirements and are also aligned with much operational practice. Furthermore, the experiment was intended to help make development decisions for the synthetic teammate, as we planned on using text-based communications when integrating the synthetic teammate with humans. Finally, given the preponderance of text-based communications in our society, the comparison of text versus voice as modes of communication is of interest in its own right.

Method

Participants

Twenty, three person teams comprised of college students and the general population of the Mesa, Arizona area voluntarily participated in one 6.5 hour session. Individuals were compensated for their participation by payment of $10.00 per person hour with each of the three team-members on the highest performing team receiving a $100.00 bonus.

The majority of the participants were males, representing 75.9% of the sample. Individuals were randomly assigned to one of three conditions: Voice Communication, Chat Communication, or Simulated Agent. The participants were also randomly assigned to teams and to role (AVO, PLO, or DEMPC, or PLO or DEMPC in the Simulated Agent

condition). All members of teams were unfamiliar with each other when they arrived for their sessions.

Equipment and Materials

The experiment took place in the CERTT Laboratory configured for the UAV-STE (described earlier). Each participant was seated at a workstation consisting of three Dell 2001 FP 20" LCD computer monitors. Two monitors were connected to an IBM PC 300PL, and a Dell Precision 220 PC for the STE. The third monitor at each workstation was connected to a Dell Precision 370 PC and was used to display the CERTT Text Chat interface during the Text Chat condition. The third monitor was not used during the Voice Communication condition. The workstation also consisted of two keyboards, one of which the participants used one for text chat communication, and the other of which participants used to enter answers into a debriefing questionnaire. Participants also used a mouse for input during the UAV task and debriefing.

Participants in the Text Chat condition communicated with each other and the experimenter using the keyboard and a custom-built text chat system designed to log speaker identity and time information. The interface was divided into 3 separate 'modules.' The 'receiver module' alerted participants with a lighted button when a message from another team member was sent. The receiver module also allowed participants to read incoming messages by pressing and holding the F10 key. Upon releasing the F10 key, the message would then be displayed in the 'storage module,' which was comprised of a window that displayed previously read messages in a list. Participants were given the ability to scroll through the messages by pressing the F7 and F8 keys to scroll down and up the list of messages. Participants sent messages with the 'transmit module.' To send messages, participants first typed their message in the transmit module window, selected the recipient using the F3, F4, and F5 keys (i.e., for the AVO, F4 corresponded to the DEMPC, F5 corresponded to the PLO, and the experimenter was always assigned to the F3 key), and then pressed F1 to send. The interface also enabled participants to select one or more recipients by clicking the appropriate receiver buttons.

Experimenters also used an identical interface to communicate with participants during missions in the Text Chat condition. In addition, the experimenter console included an interface that was used to start the chat system server, as well as log coordination events, and initiate communication glitches for situation awareness roadblocks (see below).

Participants in the Voice Communications condition communicated with each other and the experimenter using David Clark headsets and a custom-built intercom system designed to log speaker identity and time information. The intercom enabled participants to select one or more listeners by pressing push-to-talk buttons.

Two experimenters were seated in a separated adjoining room at an experimenter control station consisting of four Dell Precision 220 PCs and Dell 2001 FP 20", an IBM PC computer and Dell 2001 FP 20" monitor and four additional Dell 2001 FP 20" monitors for viewing video output and video feed from ceiling mounted Toshiba CCD cameras located behind each participant.

From the experimenter workstation, the experimenters could start and stop the missions, query participants together or individually, administer situation awareness roadblocks, log team member coordination, monitor the mission-relevant displays, select any of the computer screens to monitor using a Hall Research Technologies keyboard video mouse (KVM) matrix switch, observe team behavior through camera and audio input, and enter time-stamped observations. A Javelin Systems Quad Splitter allowed for video input from each of the four cameras to be displayed simultaneously on the monitor and was recorded on a Panasonic Omnivision VCR. In addition, a video overlay unit was used to superimpose team number, date, and real-time mission information on the video. Audio data was also recorded to the Panasonic Omnivision VCR. Furthermore, custom software recorded communication events in terms of speaker, listener, and the interval in which the push-to-talk button was depressed. A Radio Design Lab audio matrix also enabled experimenters to control the status of all lines of communication.

Custom software (seven applications connected over a local area network) ran the synthetic task and collected values of various parameters that were used as input by performance scoring software. A series of tutorials were designed in PowerPoint for training the three team members. Custom software was also developed to conduct tests on information in PowerPoint tutorials, to collect individual taskwork relatedness ratings, to collect NASA TLX and SART ratings, to administer knowledge questions, and to collect demographic and preference data at the time of debriefing.

In addition to software, some mission-support materials (i.e. rules-at-a-glance for each position, two screen shots per station corresponding to that station's computer displays, and examples of good and bad photos for the PLO) were presented on paper at the appropriate workstations. Other paper materials consisted of the consent forms, debriefing forms, and checklists (i.e. set-up, data archiving and skills training).

Procedure

The experiment consisted of one 7-hour session (see Table 1). Prior to arriving at the session, the three participants were randomly assigned to one of the three task positions: AVO, PLO or DEMPC. The team members retained these positions for the entire study. The AVO in this study was also geographically distributed from the PLO and DEMPC such that the console was located in a separate room adjacent to the other members. The AVO entered the building through a separate entrance located on the opposite side of the building, and was not allowed to have contact with the other members until debriefing.

In the session, the team members were seated at their workstations where they signed a consent form, were given a brief overview of the study and started training on the task.

During training, the PLO and DEMPC were separated by partitions (with the AVO located in a separate room). Team members studied three PowerPoint training modules at their own pace and were tested with a set of multiple-choice questions at the end of each module. If responses were incorrect experimenters provided assistance and explanations as to why their answers were incorrect and the reasoning behind the correct answers.

The PowerPoint modules for the two experimental conditions (Text Chat and Voice Communication) were identical save for the first module with regards to the training associated with the method of communication. Participants in the Text Chat condition received training on the operation of the text chat system and participants in the Voice Communication condition received training on the operation of the voice communications system.

Table 1. Experiment protocol

Consent Forms
Task Training
Mission 1
Knowledge Measures
Mission 2
Mission 3
Mission 4
NASA TLX
Knowledge Measures
Mission 5
NASA TLX
Demographics
Debriefing

After the PowerPoint phase of training, participants were then run through a short scripted communications check that lasted 10 minutes and served to allow participants to become familiar with using the CERTT Text Chat system. In the Voice Communication condition, the activity allowed experimenters and participants to make certain that all involved could communicate with each other over the headsets.

Table 2. Number of targets to be photographed, per mission

Mission	*Targets*
1	11
2	12
3	11
4	12
5	20

Once all team members completed the tutorial, test questions, and communications check, a training mission was started and experimenters had participants practice the task, checking off skills that were mastered (e.g., the AVO needed to change altitude and airspeed, the PLO needed to take a good photo of a target) until all skills were mastered. Again, the experimenters assisted in cases of difficulty. Training took a total of 1 hour and 40 minutes.

After training, the partitions were removed and the team started their first 40-minute mission. All missions required the team to take reconnaissance photos of targets. However the number of targets varied from mission to mission in accordance with the introduction of situation awareness roadblocks at set times within each mission. See Table 2 for number of targets per mission. Missions were completed either at the end of a 40-minute interval or when team members believed that the mission goals had been

completed. Immediately after each mission, participants were shown their performance scores. Participants could view their team score, their individual score, and the individual scores of their teammates. The performance scores were displayed on each participant's computer and shown in comparison to the mean scores achieved by all other teams (or roles) who had participated in the experiment up to that point

After the first mission, taskwork knowledge measures were administered. The participants were separated by partitions during the knowledge sessions as well. Once the knowledge measures were completed, partitions were removed and teams began the second 40-minute mission followed by the second, third, fourth missions, NASA TLX, second knowledge session, mission 5, and a second NASA TLX. The experiment then concluded with a demographics questionnaire and debriefing.

Results

There are five sets of analyses conducted. Reported analyses include: coordinated assessment of situations by teams (CAST), communication synchronicity, knowledge, performance, and process and coordination. The analyses performed include two levels of communication mode (voice, text), 4 missions, and one high workload mission. Workload is analyzed when appropriate, where the fifth mission performed (high-load) is compared to the fourth mission performed (low-load).

Coordinated Assessment of Situations by Teams (CAST)

These analyses were still in progress at the time this report was written.

Performance

First the general findings are reported followed by the analyses that lead to these findings.

- Team performance increased with experience.
- The main effect of communication mode (text, voice) did not significantly affect team performance ($p = 0.46$).
- Load affected team performance, where team performance decreased with increased load, as expected.
- Across the three roles, PLO was the only role to demonstrate an effect of communication mode on performance, with PLO participants in the voice communications condition performing better than PLO participants in the text-based communications condition.

Team performance was measured using a composite score based on the result of mission variables including time each individual spent in an alarm state, time each individual spent in a warning state, rate with which critical waypoints were acquired, and the rate with which targets were successfully photographed. Penalty points for each of these components were weighted *a priori* in accord with importance to the task and subtracted from a maximum score of 1000. Team performance data were collected for each of the seven missions.

Each individual role within a team (AVO, PLO and DEMPC) also had a composite score based on various mission variables including time spent in alarm or warning state as well as variables that were unique to that role. Penalty points for each of the components were weighted *a priori* in accord with importance to the task and subtracted from a maximum score of 1000. The most important components for the AVO were time spent in alarm state and course deviations, for the DEMPC they were critical waypoints missed and route planning errors, and for the PLO, duplicate good photos, time spent in an alarm state, and number of bad photos were the most important components. *Individual performance* data for a role were collected for each of the seven missions.

This team performance measure has been used in previous CERTT studies and was modified in the last effort (Cooke, et al., 2004) in order to take into account workload differences in scenarios. For example, the new team performance metric, which is based on rate of performance, does not penalize teams for photographing a smaller proportion of targets in the high workload missions (e.g., 12 out of 20 targets) despite the improvement from the low workload missions (e.g., 9 out of 9 targets). Appendix A shows the weighting scheme used for each component of the team and individual role performance metrics.

Team Performance

The team performance score is calculated from several sub-components. These components are alarms, warnings, fuel, film, route sequence violation, critical waypoints per minute, and missed/slow photos. Alarms and warnings are a measure of the percentage of mission time that team members were in alarm or warning states. These percentages are cumulative across team members. Critical waypoints per minute refers to the number of target waypoints and restricted operating zone entries and exits that the team visited per minute. Photo rateis a measure of how many good target photos per minute were obtained by the PLO over the mission.

Team performance was analyzed using a 2 (text, voice) x 4 (mission) mixed ANOVA. Each communication condition (text, voice) had 10 teams. The analysis results indicate a main effect of mission $F(3, 54) = 9.447, p < .001$. There were no significant effects of communication condition, $F(1, 18) = 0.57, p < 0.46$, although the voice communication teams consistently had higher performance scores across all missions. The following shows the performance scores across missions for each communication condition.

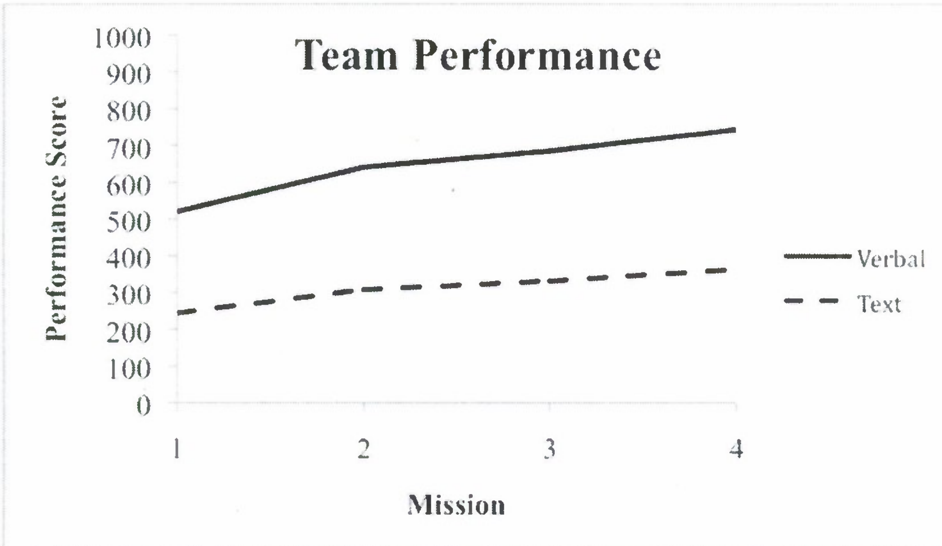

Figure 4. Team performance means for each mission.

LSD pair-wise comparisons showed that team performance improved over the course of the first four missions, with significant gains between the first two missions (p = .005) and between the second and fourth missions (p = .015).

A 2 (text, voice) x 2 (baseline workload, high workload) mixed ANOVA was performed to assess the effect of workload on team performance. The results indicate a main effect of workload $F(1,18) = 11.47$, $p = .003$ (see Figure 5). There was not a main effect of communication, $F(1,18) = 1.27$, $p = 0.274$, nor was there a communication x workload interaction, $F(1,18) = 0.848$, $p = 0.369$.

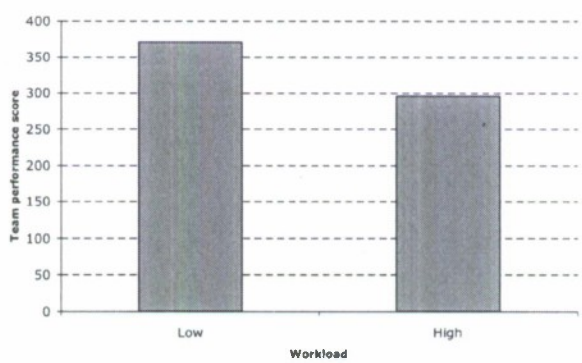

Figure 5. Workload effect on team performance.

AVO Performance

The AVO's performance score is based on four penalty scores: alarms, warnings, fuel, course deviation, and route sequence. The alarm and warning penalties are based on the amount of time that the AVO spends in alarm and warning states. Course deviation refers to how well the AVO stays on the course needed to get to each waypoint, while route sequence refers to how well the AVO follows the planned route sent by the DEMPC.

A 2 (text, voice) x 4(mission) mixed ANOVA was performed to assess individual performance for the AVO role. The results for this test revealed a main effect of mission $F(3, 18) = 5.592, p = .002$.

LSD pair-wise comparisons showed that AVO performance improved between the second and third missions ($p = .02$) but then leveled off. The scores across mission for each communication condition can be seen in Figure 6.

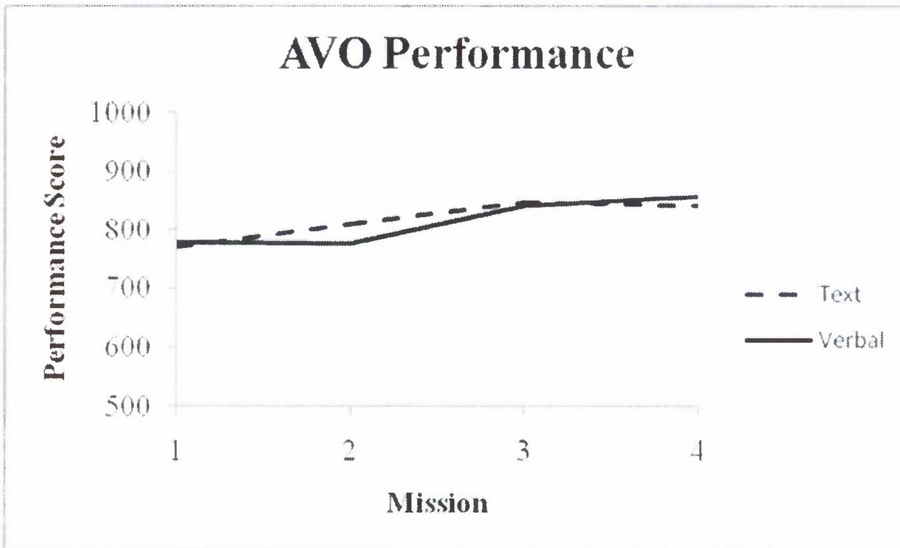

Figure 6. AVO performance scores across missions.

AVO Workload Analysis

A 2 (text, voice) x 2 (baseline workload, high workload) mixed ANOVA was performed to assess the effect of workload on AVO performance. The results indicate a main effect of workload $F(1,18) = 6.796, p = .018$. (see Figure 7).

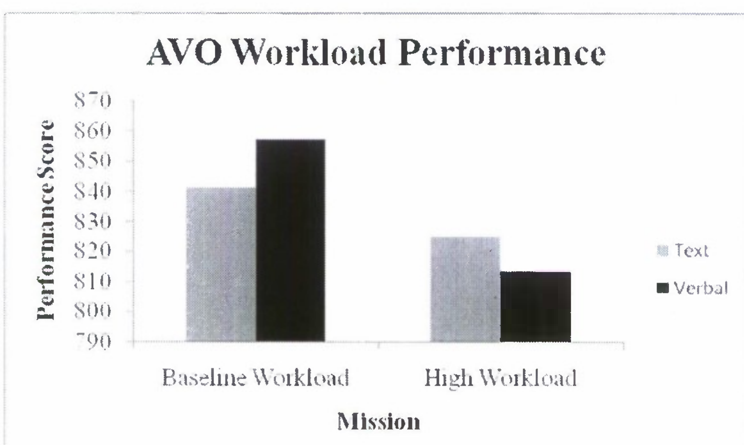

Figure 7. AVO workload performance.

DEMPC Performance

The DEMPC's performance score is based on five penalties: alarms, warnings, missed critical waypoints not planned, alarm waypoints, and route sequence planning. Missed critical waypoints not planned is a penalty for waypoints that should have been visited but were missed because they were never added to the route plan. Alarm waypoints is a penalty for visiting hazardous waypoints. Route sequence planning refers to how well the DEMPC followed the rules regarding priority targets and restricted operating zone entrances and exits.

A 2 (text,voice) x [4 (mission)] mixed ANOVA was used to examine DEMPC performance. The Greenhouse-Geisser correction is reported because the sphericity assumption was violated. Analyses showed a main effect of mission $F(2.121, 18) = 8.501, p = .001$.

LSD Pair-wise comparisons indicated a significant improvement between the first and second missions ($p = .07$) and significant improvement between the first and third missions ($p = .003$). Performance appears to level off after the third mission. Figure 8 shows the DEMPCs' performance across missions.

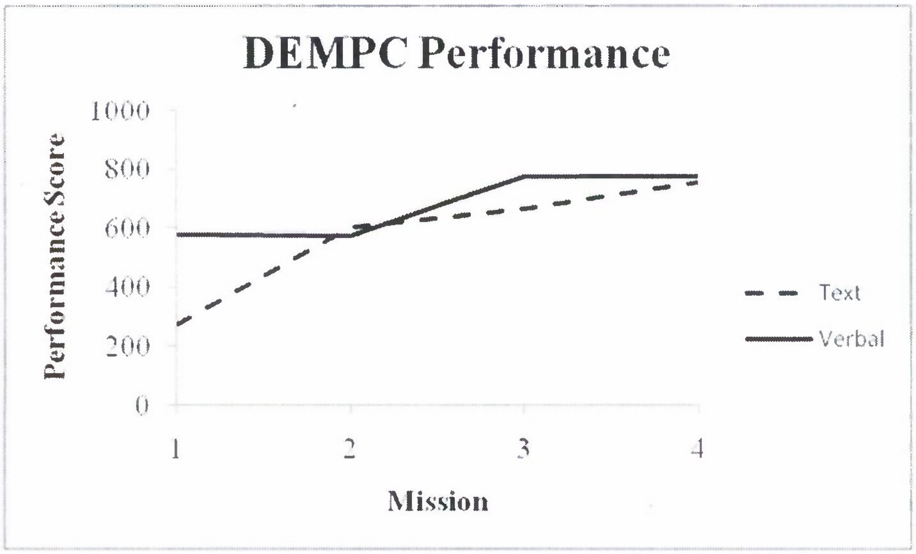

Figure 8. DEMPC performance across missions.

A 2 (text, voice) x 2 (baseline workload, high workload) mixed ANOVA was performed to asses the effect of workload on DEMPC performance. The results indicate a main effect of workload $F(1,18) = 57.651, p < .001$ (see Figure 9).

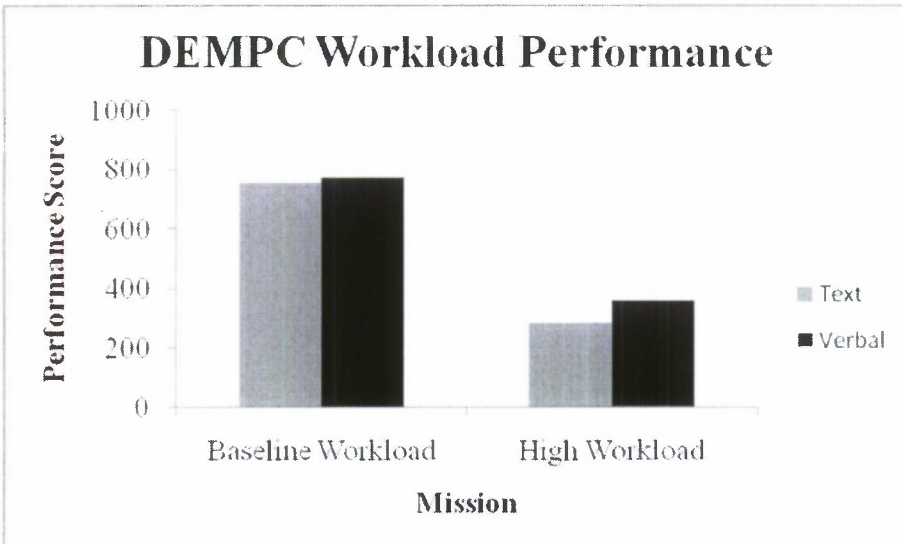

Figure 9. DEMPC workload performance.

PLO Performance

The PLO's score is also based on five penalties: alarms, warnings, duplicate photos, bad photos, and missed/slow photos. Alarm and warning penalties are calculated the same way as for the AVO. Duplicate photos refers to the number of times the PLO took a photo of a target that had already been successfully photographed. Bad photos is the number of unsuccessful photo attempts. Photo rate is a measure of how many good target photos per minute were obtained by the PLO over the mission.

PLO performance scores were analyzed using a 2 (text,voice) x [4 (mission)] mixed ANOVA. There was one outlier that was excluded from the analyses because his/her mean performance score was greater than 3 standard deviations from the mean PLO performance score. After removing the outlier, there were 10 PLOs in the chat communication condition and nine PLOs in the voice communication condition.

Results of the mixed ANOVA indicated a main effect of condition and $F(1, 17) = 9.95$, $p = .006$. Specifically, PLOs in the voice communication condition performed better than PLOs in the text chat condition. In addition, the analysis revealed a main effect of mission $F(3, 17) = 4.076$, $p = .011$. Figure 10 shows PLO performance for each communication condition over the four missions.

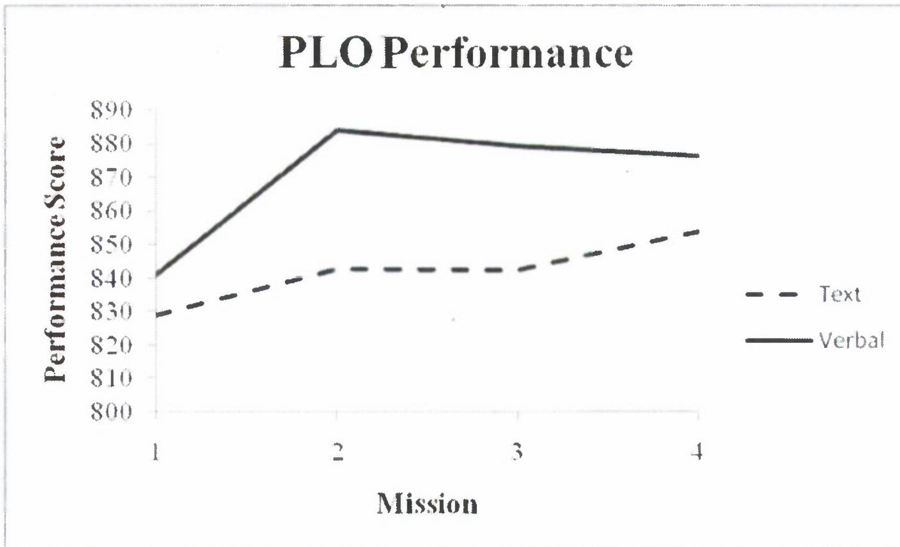

Figure 10. Mean PLO performance across missions.

A 2 (text, voice) x 2 (baseline workload, high workload) mixed ANOVA was performed to assess the effect of workload on PLO performance. The results indicate a main effect of workload $F(1,17) = 7.066$, $p = .017$, and a main effect of communication condition $F(1,17) = 3.882$, $p = .065$ (see Figure 11).

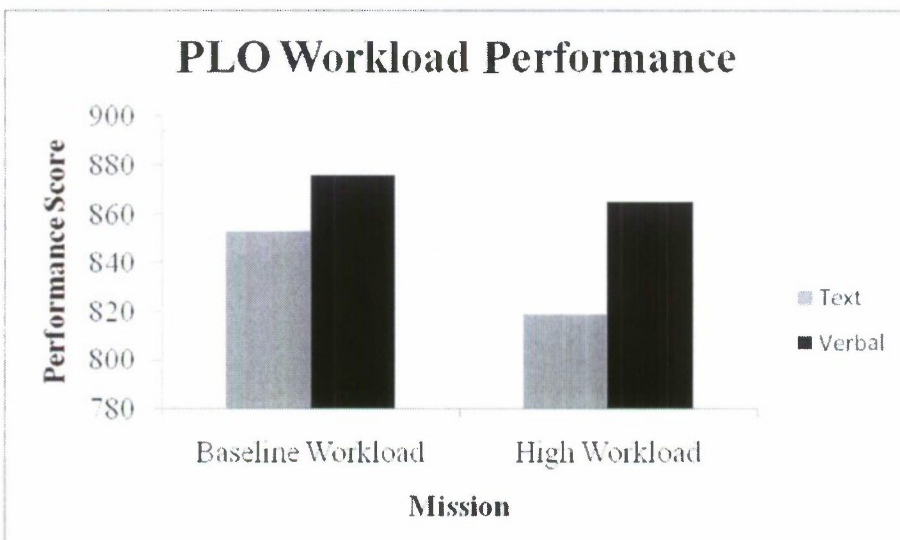

Figure 11. PLO workload performance.

Subjective Workload Ratings

The NASA TLX questionnaire was used to determine the mental, physical, and temporal demand that participants experienced across different missions. The questionnaire was also used to determine participants' degree of efficacy. These dimensions were rated on a scale of 0-100 by each participant and then multiplied by a weighted value. The products were then summed in order to arrive at a total score (see Hart & Staveland, 1988).

Teams 1 and 2 do not have any TLX data. Team 4 mission 9 AVO data was unreadable for performance and total score. The PLO for team 18 was a performance outlier and was excluded from workload analyses. Consequently, there were 18 (17 for performance and total score analyses) AVO, 17 PLO, and 18 DEMPC TLX scores analyzed.

Team positions were averaged across mission (fourth low-load mission and high-load mission) for each participant. The resulting scores were tested for normality in SPSS.

Mental Demand

Mental demand on participants was analyzed using a 2 (text, voice) x 3 (AVO, PLO, DEMPC) x 2 (mission) mixed ANOVA. Results indicated that there was a main effect of role, $F(2,47) = 56.876$, $p < 0.001$. Planned pair-wise comparisons using the LSD correction (i.e., no correction for family-wise error) revealed that DEMPCs experienced greater mental demand than both AVOs ($p < 0.001$) and PLOs ($p < 0.001$).

There was not a significant main effect of mission, $F(1,47) = 1.349$, $p = .251$, or communication condition, $F(1,47) = .563$, $p = .457$. There were no interactions between mission and communication condition, $F(1,47) = .869$, $p = .356$, or between mission and role, $F(2,47) = .872$, $p = .425$. The interaction between mission, communication condition, and role was also not significant, $F(2,47) = .75$, $p = .478$. There was no interaction between communication condition and role, $F(2,47) = .280$, $p = .757$. Figure 12 shows mean mental demand ratings for each role across the low-load and high-load missions.

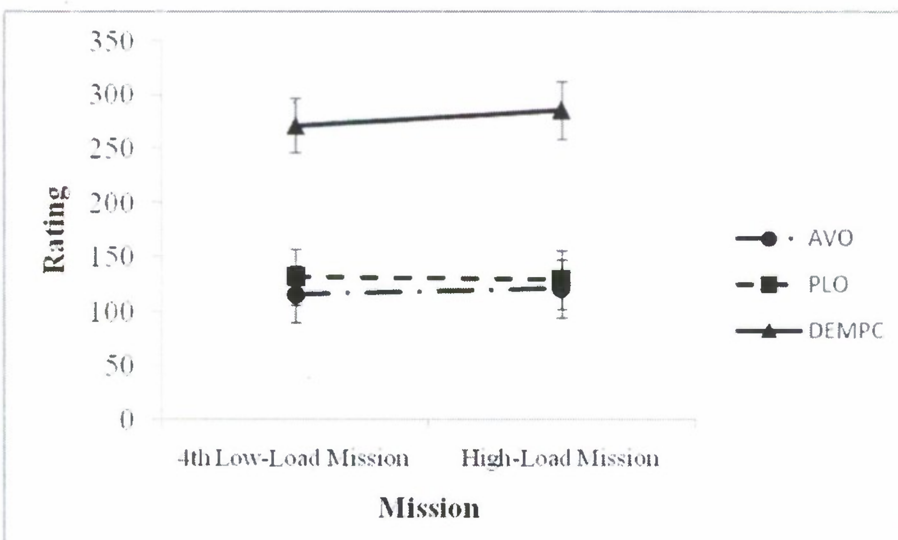

Figure 12. Mean mental demand ratings for each role across mission. Error bars represent 95% confidence intervals.

Physical Demand

A 2 (text, voice) x 3 (AVO, PLO, DEMPC) x 2 (mission) mixed ANOVA was used to analyze the physical demand experienced by participants. Results showed a main effect of role, $F(2,47) = 9.203$, $p < 0.001$. Planned pair-wise comparisons using the LSD

correction (i.e., no correction for family-wise error) revealed that AVOs felt more physical demand than PLOs ($p = .002$) and DEMPCs ($p < 0.001$).

There was not a significant main effect of mission, $F(1,47) = 1.087, p = .302$, or communication condition, $F(1,47) = .470, p = .496$. There were no interactions between mission and communication condition, $F(1,47) = .652, p = .424$, or between mission and role, $F(2,47) = .146, p = .865$. The interaction between mission, communication condition, and role was also not significant, $F(2,47) = 1.277, p = .288$. There was no interaction between communication condition and role, $F(2,47) = 1.522, p = .229$. Figure 13 shows mean physical demand ratings for each role across the low-load and high-load missions.

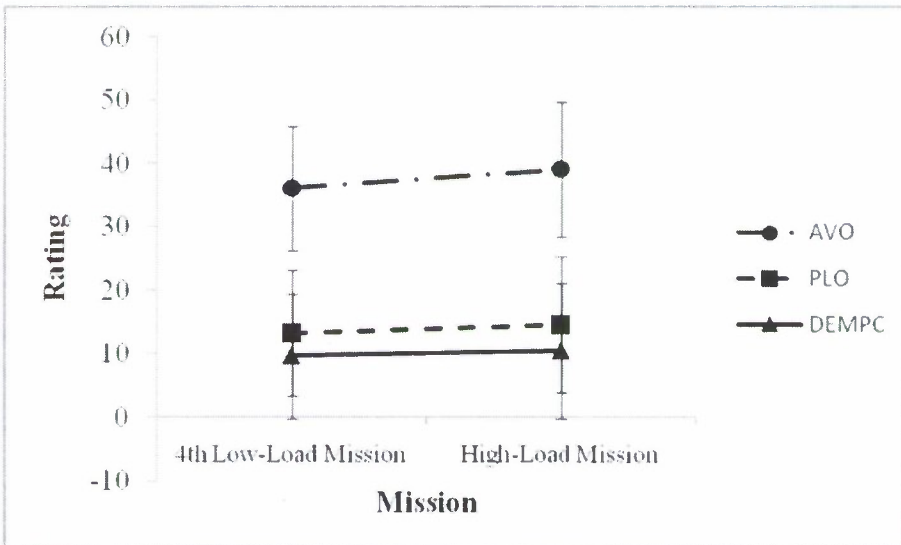

Figure 13. Mean physical demand ratings for each role across mission. Error bars represent 95% confidence intervals.

Temporal Demand

A 2 (text, voice) x 3 (AVO, PLO, DEMPC), x 2 (mission) mixed ANOVA was used to analyze the time pressure felt by participants. A significant mission x communication condition x role interaction was found, $F(2,47) = 2.45, p = .097$. AVOs in the text condition showed an increase in temporal demand from the fourth low-load mission (M = 185.71, SD = 16.04) to the high-load mission (M = 206.48, SD = 18.5), while AVOs in the voice condition did not show an increase in temporal demand from the low-load mission (M = 181.86, SD = 16.04) to the high-load mission (M = 184.23, SD = 18.5). PLOs in the text condition demonstrated an increase in temporal demand from the low-load mission (M = 148.61, SD = 16.04) to the high-load mission (M = 164.44, SD = 18.5), but PLOs in the voice condition showed a decrease in temporal demand from the low-load mission (M = 160.63, SD = 17.01) to the high-load mission (M = 125.94, SD = 19.62). The DEMPCs in each communication condition showed stable temporal demand ratings between the low-workload mission and the high-workload mission. Means for the text DEMPCs were 83.85, SD = 16.04 (low-load) and 81.51, SD = 18.5 (high-load). DEMPCs in the voice condition had means of 79.56, SD = 16.04 (low-load) and 77.09,

SD = 18.5 (high-load). Figure 14 shows the mean temporal demand ratings for each role in each communication condition.

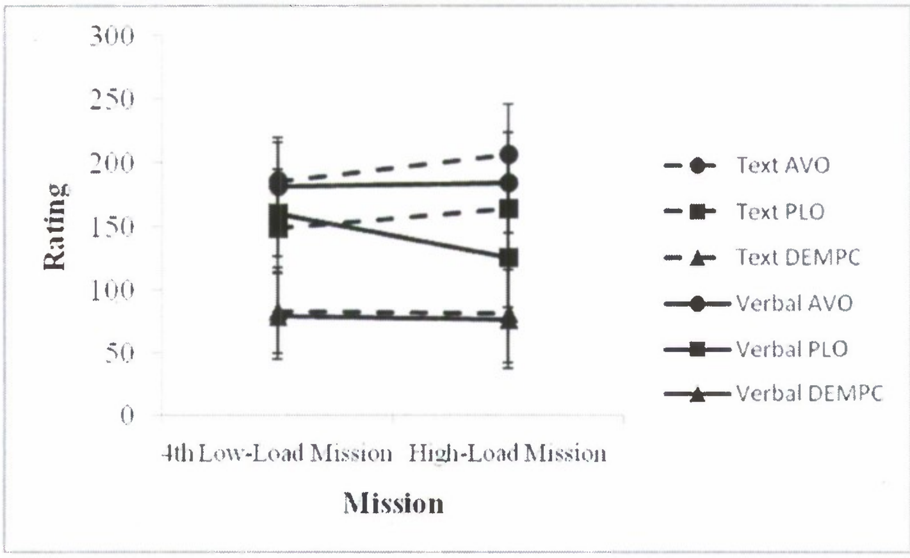

Figure 14. Mean temporal demand ratings for each role in communication condition. Error bars represent 95% confidence intervals.

The analysis further revealed a significant mission x condition interaction, $F(1,47) = 6.071$, $p = 0.017$. Participants in the text condition experienced an increase in temporal demand from the fourth low-load mission (M = 139.39, SD = 9.26) to the high-load mission (M = 150.81, SD = 10.68), but participants in the voice condition experienced a decrease in temporal pressure from the low-load mission (M = 140.68, SD = 9.45) to the high-load mission (M = 129.09, SD = 10.9). Figure 15 shows the mean temporal demand ratings in each communication condition across the fourth low-load mission and the high-load mission.

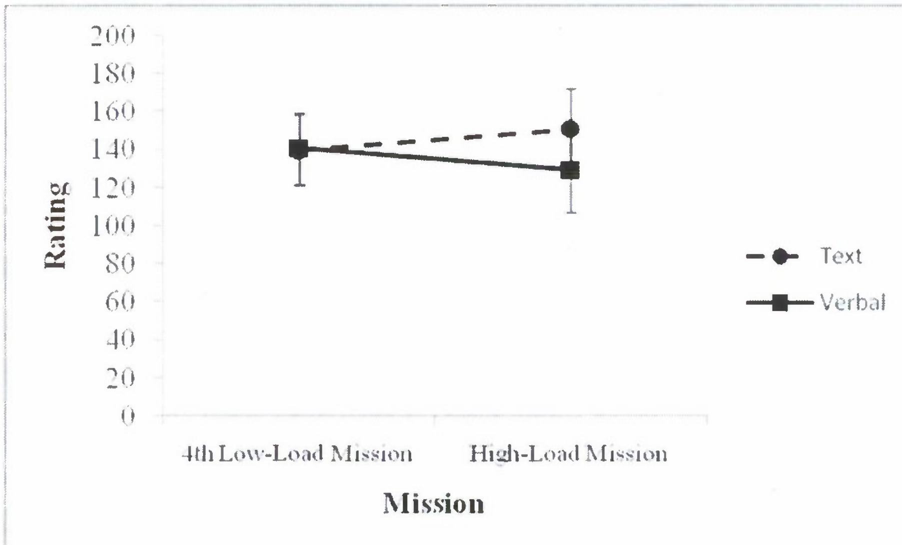

Figure 15. Mean temporal demand ratings for each communication condition across mission. Error bars represent 95% confidence intervals.

There was a main effect of role, $F(2,47) = 22.753$, $p < 0.001$. LSD pair-wise comparisons showed that all roles experienced significantly different amounts of time pressure, with the AVO experiencing the most pressure, followed by the PLO, with the DEMPC experiencing the least amount of time pressure. Figure 16 shows the mean temporal demand ratings of each role across the low-load and high-load missions.

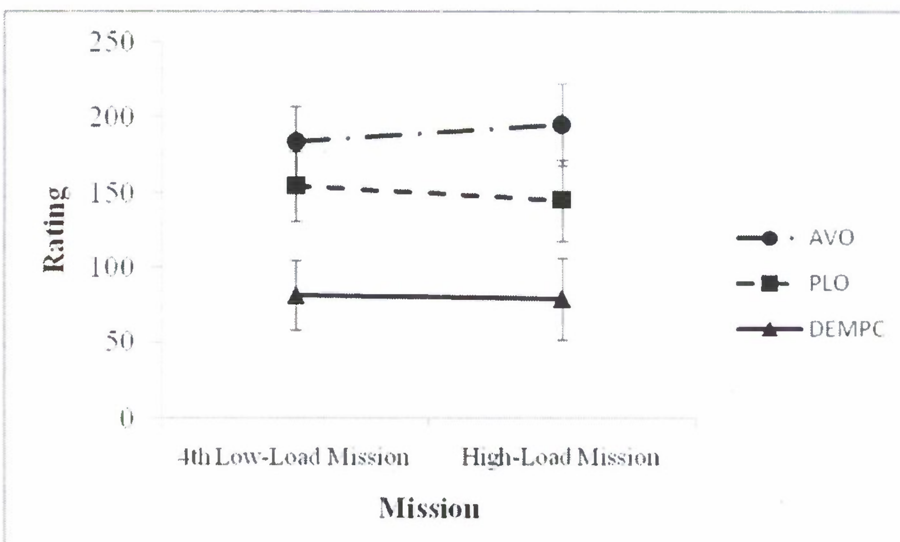

Figure 16. Mean temporal demand ratings of each role across mission. Error bars represent 95% confidence intervals.

There was not a significant main effect of mission, $F(1,47) < 0.001$, $p = .985$, or communication condition, $F(1,47) = .573$, $p = .453$. There were no interactions between mission and role, $F(2,47) = 1.741$, $p = .186$. There was no interaction between communication condition and role, $F(2,47) = .048$, $p = .953$.

Performance

There were no significant effects of mission, $F(1,46) = .621$, $p = .435$, communication condition, $F(1,46) = .151$, $p = .700$, or role, $F(2,46) = 2.197$, $p = .123$. There was also no interaction between mission, communication condition, and role, $F(2,46) = .291$, $p = .749$. No interactions were found between mission and communication condition, $F(1,46) = .179$, $p = .675$, or between mission and role, $F(2,46) = .403$, $p = .670$. There was no interaction between communication condition and role, $F(2,46) = .667$, $p = .518$.

Total Score

There were no significant effects of mission, $F(1,46) = 1.913$, $p = .173$, communication condition, $F(1,46) = .059$, $p = .809$, or role, $F(2,46) = .328$, $p = .722$. There was also no interaction between mission, communication condition, and role, $F(2,46) = 1.863$, $p = .167$. No interactions were found between mission and communication condition, $F(1,46) = .734$, $p = .396$, or between mission and role, $F(2,46) = .983$, $p = .382$. There was no interaction between communication condition and role, $F(2,46) = .109$, $p = .897$.

Summary

The DEMPC perceives the greatest amount of mental demand, while the AVO experiences the greatest physical and temporal demands from the task. Furthermore, the text-communication condition AVOs and PLOs experience greater temporal demand as workload increases, while the voice-communication condition AVOs and PLOs experience the same and less temporal demand, respectively. The DEMPCs in each communication condition maintain stable levels of temporal demand as workload increases.

Communication Synchronicity

First the general findings are reported followed by the analyses that lead to these findings.

- Communication lag times (received time minus sent time) for text-based communications were > 0, demonstrating communication asynchrony.
- Lag times interacted with role (i.e., PLO, DEMPC, AVO) and experience (missions 1-4), where AVO and DEMPC reduced their lag times and the PLO's increased with experience.
- Lag times interacted with role and cognitive load (high and low), where lag times for the AVO decreased from low to high load and PLO and DEMPC lag times increased from low to high load.

The two communication modes, voice and text, provide significantly different forms of communications. The obvious differences include visual (text) as opposed to auditory (voice) inputs and manual (text) versus voice (voice) outputs. However, the two are also different in how rapid communications are received. The receiver interprets voice communications as the communication is sent (i.e., the communication transmission and receipt are *synchronous*). The receiver of text-based communications can either interpret

a message as soon as it is sent (i.e., synchronous) or sometime after it is sent (i.e., *asynchronous*).

A 3 x 4 mixed ANOVA was performed on the difference between the time a text message was sent and received, in seconds (i.e., com-lag), to determine if different CERTT task positions took longer to receive sent messages across the five task missions in the text condition. Mission violated the sphericity assumption; hence the Greenhouse-Geisser correction was used where applicable. One outlier was removed from the analysis because his/her mean time-lag was greater than three standard deviations from the mean time lag. There was a significant mission x task position interaction $F(4.25, 53.08) = 2.63$; $p = 0.042$, $MSE = 92.66$; $M_{AVO} = 6.84$, $M_{PLO} = 12.71$, $M_{DEMPC} = 11.97$; $M_{Mission\text{-}1} = 14.12$, $M_{Mission\text{-}2} = 9.71$, $M_{Mission\text{-}3} = 9.53$, $M_{Mission\text{-}4} = 9.16$ (see Figure 17).

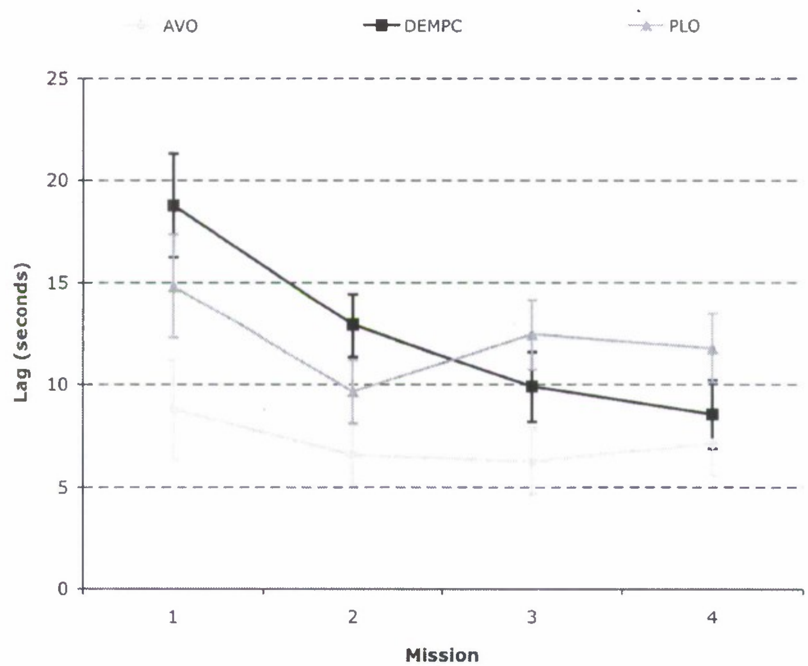

Figure 17. Communications lag analysis

The results indicate that the lag time of communication receptions was a function of mission and teammate position. Furthermore, the results demonstrated that the text-based communication condition functioned as an asynchronous communication platform across teammates and missions.

To determine how high workload affected communication synchronicity, the communication lag times from the final low-load mission (i.e., mission 4) were compared to lag times from the high-load mission (i.e., mission 5). A 2 (load) x 3 (role) mixed ANOVA was performed. There was a significant load (high, low) x role (PLO, AVO, DEMPC) interaction, $F(2, 26) = 3.149$, $p = 0.60$, $MSE = 26.69$ (see Figure 18).

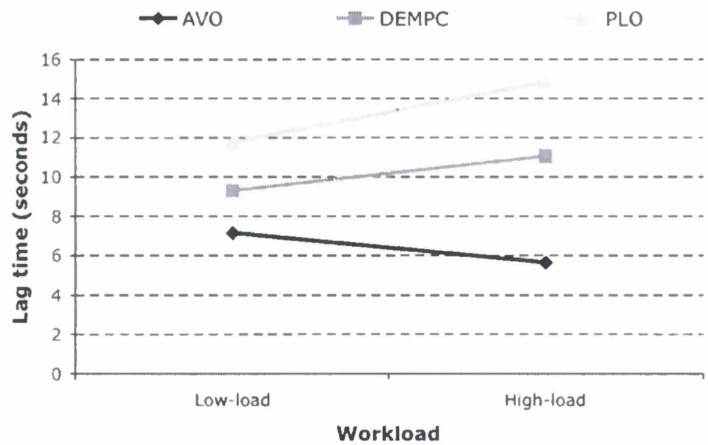

Figure 18. Workload effects on communication lag times by teammate role

Taskwork Knowledge

Taskwork knowledge was assessed through a rating task (see Appendix B). The taskwork ratings consisted of eleven task related terms: altitude, focus, zoom, effective radius, ROZ entry, target, airspeed, shutter speed, fuel, mission time, and photos. These task-related terms formed 55 concept pairs, which were presented in one direction only, one pair at a time. Pair order was randomized and order within pairs was counterbalanced across participants.

Team members made relatedness ratings of the 55 concept pairs on a six-point scale that ranged from unrelated to highly-related. By submitting these ratings to Knowledge Network Organization Tool (KNOT), using parameters r = infinity and q = n-1, an individual Pathfinder network (Schvaneveldt, 1990) was derived for each of the team members. These networks reduce and represent the rating data in a graph structure with concept nodes standing for terms and links standing for associations between terms. The individual taskwork networks were scored against a key representing overall knowledge, and against role-specific keys. In this way, measures of "role" or "positional" accuracy, as well as "interpositional" accuracy could be determined. The referent networks were based on data from the highest scoring individuals or teams in our previous studies.

The accuracy of an individual's knowledge was determined by comparing each individual network to empirical referents associated with knowledge relevant to the respective roles and overall knowledge. Network similarities were computed that ranged from 0 to 1 and represented the proportion of shared links between the two networks (based on the Pathfinder similarity metric).

Using this similarity metric, three accuracy values were computed for each team member. Overall accuracy is the similarity between the individual network and the overall knowledge referent. Positional (role) accuracy is the similarity between the individual's network and the referent network associated with that individual's role. Interpositional accuracy is the average of the similarity between the individual's network and the referent networks of the two other roles. These three accuracy values were averaged

across all team members to give a final overall, positional and interpositional accuracy score for each team. It should be noted that prior to averaging similarity values to calculate positional and interpositional accuracy scores for the team, positional and interpositional scores for each team member were standardized, as team positional and interpositional accuracy scores are made up of individual scores based on different referents.

Intrateam similarity was scored on the same scale as accuracy and ranged from 0 to 1. An individual's network was compared to another team member's network and assigned a similarity value. This was done until all three team members had been compared to one another (i.e. AVOPLO, AVO-DEMPC, and PLO-DEMPC). Intrateam similarity was computed by averaging the three similarity values measured using the proportion of shared links for all intrateam pairs of two individual networks (i.e. the mean of the three pairwise similarity values across the three networks).

First the general findings are reported followed by the analyses that lead to these findings.

- For *interpositional knowledge accuracy*, teams showed a significant increase in Taskwork Knowledge from Session 1 to Session 2.
- For *intrateam similarity,* teams showed a significant increase in Taskwork Knowledge from Session 1 to Session 2
- The increases in Taskwork Knowledge are attributable increased communication and knowledge gathering about other team-members' roles.

Taskwork Overall Accuracy

An examination of Q-Q plots showed that the dependent variable was approximately normally distributed. An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in overall accuracy, $F(1, 15) = .028, p = .87$.

A repeated measures ANOVA investigated whether there was a change in taskwork overall accuracy for all teams (regardless of Treatment) from Knowledge Session 1 to 2. The analysis revealed that all teams in general, did not significantly improve in overall accuracy from Session 1 to Session 2, $F(1, 15) = 1.95, p = .183$.

Taskwork Positional Knowledge

An examination of Q-Q plots showed that the dependent variable was approximately normally distributed. An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in overall positional knowledge, $F(1, 15) = 2.18, p = .160$.

A repeated measures ANOVA investigated whether there was a change in taskwork overall accuracy for all teams (regardless of Treatment) from Knowledge Session 1 to 2. The analysis revealed that all teams in general, did not significantly improve in overall accuracy from Session 1 to Session 2, $F(1, 15) = 2.02, p = .176$.

Taskwork Interpositional Knowledge

An examination of Q-Q plots showed that the dependent variable was approximately normally distributed. An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in interpositional knowledge, $F(1, 15) = .134, p = .719$.

A repeated measures ANOVA investigated whether there was a change in taskwork interpositional accuracy for all teams (regardless of Treatment) from Knowledge Session 1 to 2. The analysis revealed that teams in general, significantly improved in overall accuracy from Session 1 to Session 2, $F(1, 15) = 9.04, p = .009$.

However, analyses separately comparing Knowledge Sessions for the different Communication Modes revealed that both teams in the Text and Voice conditions significantly increased $F(1, 7) = 4.83, p = .06$, and $F(1, 8) = 4.11, p = .077$ respectively.

Taskwork Intrateam Similarity

An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in intrateam similarity, $F(1, 15) = .151, p = .703$.

A repeated measures ANOVA investigated whether there was a change in taskwork intrateam similarity for all teams (regardless of Treatment) from Knowledge Session 1 to 2. The analysis revealed that teams in general, significantly improved in similarity from Session 1 to Session 2, $F(1, 15) = 8.356, p = .011$.

However, analyses separately comparing Knowledge Sessions for the different Communication Modes revealed that both teams in the Text and Voice conditions significantly increased, $F(1, 7) = 4.11, p = .082$, and $F(1, 8) = 4.21, p = .074$ respectively.

Correlations between taskwork knowledge and team performance

Analysis of taskwork knowledge revealed significant findings for interpositional knowledge accuracy and intrateam similarity. To observe the relationship between this knowledge measure and team performance, the interpositional knowledge accuracy and intrateam similarity scores obtained during Knowledge Session 1 were correlated with team performance scores obtained during Mission 4. The results of performed correlations are presented in Table 3. The lack of any significant correlations indicates that taskwork positional accuracy and performance measures are not linearly related.

Table 3. Correlations between teamwork interpositional knowledge,and team performance.

	Team performance score during Mission 4
Taskwork *interpositional knowledge accuracy* score during Knowledge Session 1	-.317
Taskwork *intrateam similarity* score during Knowledge Session 1	.424

To observe the relationship between all dependent variables (overall accuracy, positional knowledge, interpositional knowledge, and intrateam similarity) with regard to the first and second knowledge sessions, correlations were performed. The correlations are shown in Table 4.

Table 4. Correlations between taskwork measures comparing Session 1 to Session 2.

	Session 1 and Session 2 Correlation
Overall accuracy	-.082
Positional knowledge	.388
Interpostional knowledge	.183
Intrateam similarity	.744*

Session 1 Intrateam similarity was found to be significantly correlated with its Session 2 counterpart at the $p = .01$ level. Following this finding, a MANOVA using all Knowledge Session 1 and Knowledge Session 2 taskwork measures as dependent variables with Communication Mode as the fixed factor was performed. The MANOVA however, revealed no significant results.

Teamwork Knowledge

Teamwork knowledge was assessed using a teamwork questionnaire (see Appendix C). The teamwork questionnaire consisted of a scenario in which each individual participant was required to indicate which of sixteen specific communications were absolutely necessary in order to achieve the scenario goal. To calculate each individual's overall accuracy, the responses were compared to an answer key, which classified each of the 16 communications into one of the following categories: (1) the communication is NEVER absolutely necessary to complete the scenario goal; (2) the communication could POSSIBLY be necessary to complete the scenario goal (e.g., as considered by novices); or (3) the communication is ALWAYS absolutely necessary to complete the scenario goal. Each communication was worth 2 points, which yielded a maximum of 32 points possible per team member. Participants either checked each communication, indicating that it was absolutely necessary to complete the scenario goal or left it blank, indicating that it wasn't absolutely necessary. The table below illustrates how the questionnaires were scored. A perfect score was achieved by only checking those communications that were ALWAYS absolutely necessary and leaving all other communications blank. Team overall knowledge was the mean of the three team members' overall accuracy scores.

Using the same scoring scheme, individual team member responses to the teamwork questionnaire were also scored against role-specific keys. In particular, "role" or "positional" accuracy, as well as "interpositional" accuracy (i.e., interpositional knowledge or knowledge of roles other than his or her own) was determined. Role or positional knowledge accuracy was determined by comparing each individual's responses to the role-specific key. To score positional knowledge accuracy, each role-specific key was used to compare each individual's responses to the subset of the items on the

questionnaire specific to his/her role. For example, the key for AVO positional knowledge did not take into consideration five items on the questionnaire that asked about communications between PLO and DEMPC. Therefore, the maximum score for AVO positional knowledge accuracy was 22 (i.e., 11 questionnaire items worth 2 points each). The maximum scores for PLO and DEMPC positional knowledge accuracy were 20 and 22, respectively. Scores were converted into proportion of points and proportions were averaged across the three team members to derive a positional accuracy score for the team.

For each role, interpositional knowledge was scored against those items on each key not used in scoring positional knowledge. For example, the accuracy of AVO's responses on the teamwork questionnaire to those 5 items involving communications between the PLO and DEMPC constituted his/her score for interpositional knowledge. Since each response is worth 2 points, the AVO interpositional knowledge maximum is 10. The maximum scores for PLO and DEMPC interpositional knowledge accuracy scores were 12 and 10, respectively. Scores were converted into proportion of points and proportions were averaged across the three team members to derive an interpositional accuracy score for the team.

Intra-team similarity was also computed by comparing responses from all 3 participants and assigning a point to every response that all the team members had in common. A maximum of 16 points were possible where a higher score indicates that more of the team members' responses were identical.

First the general findings are reported followed by the analyses that lead to these findings.

- Data for all teamwork measures were homogeneous and approximately normally distributed.
- For *overall accuracy*, Text Communication mode teams showed a significant decrease in Teamwork Knowledge from Session 1 to Session 2. This decrease may be due to limitations in the amount of communication possible in the Text Chat environment as well as the fact that the AVO was not co-located.
- For *interpositional knowledge accuracy*, Text Communication mode teams showed a significant decrease in Teamwork Knowledge from Session 1 to Session 2. This decrease may also be due to limitations in the amount of communication possible in the Text Chat environment as well as the fact that the AVO was not co-located.

Teamwork Overall Accuracy

An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in overall accuracy, $F(1, 15) = .177$, $p = .68$. A repeated measures ANOVA investigated whether there was a change in teamwork overall accuracy for all teams (regardless of Communication Mode) from Knowledge Session 1 to 2. The analysis revealed that teams in general, significantly *worsened* in overall accuracy from Knowledge Session 1 to Knowledge Session 2, $F(1, 15) = 3.31$, $p = .09$.

Teamwork Positional Knowledge Accuracy

An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in positional accuracy, $F(1, 15) = .000$, $p = .986$. A repeated measures ANOVA investigated whether there was a change in teamwork positional accuracy for all teams (regardless of Communication Mode) from Knowledge Session 1 to 2. The analysis revealed that teams in general, did not change in positional accuracy from Knowledge Session 1 to Knowledge Session 2, $F(1, 15) = 2.87$, $p = .11$.

However, analyses separately comparing Knowledge Sessions for the different Communication Modes revealed that teams in the Text condition significantly decreased $F(1, 7) = 5.47$, $p = .05$, while Voice teams did not show a change, $F(1, 8) = .001$, $p = .97$.

Teamwork Interpositional Knowledge

An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in interpositional accuracy, $F(1, 15) = .97$, $p = .34$. A repeated measures ANOVA investigated whether there was a change in teamwork overall accuracy for all teams (regardless of Communication Mode) from Knowledge Session 1 to 2. The analysis revealed that teams in general, significantly *worsened* in interpositional accuracy from Knowledge Session 1 to Knowledge Session 2, $F(1, 15) = 3.22$, $p = .09$.

Analyses separately comparing Knowledge Sessions for the different Communication Modes revealed that teams in the Text condition significantly decreased $F(1, 7) = 5.37$, $p = .05$, while Voice teams did not show a change, $F(1, 8) = .157$, $p = .70$.

Teamwork Intrateam Similarity

An analysis of the between-subjects effects revealed no main effects of Communication Mode indicating that teams performed similarly in intra-team similarity, $F(1, 15) = .229$, $p = .639$. A repeated measures ANOVA investigated whether there was a change in teamwork intra-team similarity for all teams (regardless of Treatment) from Knowledge Session 1 to 2. The analysis revealed that teams in general, did not significantly improve in intra-team similarity from Session 1 to Session 2, $F(1, 15) = 1.044$, $p = .323$.

Correlations between teamwork knowledge measure and team performance

Analysis of teamwork knowledge revealed significant findings for overall accuracy, and interpositional accuracy. To observe the relationship between this knowledge measure and team performance, interpostional knowledge accuracy scores obtained during Knowledge Session 1 were correlated with team performance scores obtained during Mission 4 (performance asymptote). The results of performed correlations are presented in Table 5. The lack of any significant correlations indicates that teamwork and performance measures are not linearly related.

Table 5. Correlations between teamwork overall aeeuraey, interpositional knowledge,and team performanee.

	Team performance score during fourth mission
Teamwork *interpositional knowledge accuracy* score during Knowledge session 1	.012
Teamwork *overall knowledge accuracy* score during Knowledge session 1	.028

To observe the relationship between all dependent variables (overall accuracy, positional knowledge, interpositional knowledge, and intrateam similarity) with regard to the first and second knowledge sessions, correlations were performed. The correlations are shown in Table 6.

Table 6. Correlations between taskwork measures comparing Session 1 to Session 2.

	Session 1 and Session 2 Correlation
Overall accuracy	.475
Positional knowledge	.684*
Interpostional knowledge	.420
Intrateam similarity	.421

Session 1 Positional accuracy was found to be significantly correlated with its Session 2 counterpart at the $p = .01$ level. Following this finding, a MANOVA using all Knowledge Session 1 and Knowledge Session 2 teamwork measures as dependent variables with Communication Mode as the fixed factor was performed. The MANOVA however, revealed no significant results.

Team Process & Coordination

Team Coordination Log

The team coordination logger is a custom-developed software tool that allows for the recording and time stamping of team coordination events in the CERTT Lab UAV-STE. This measure is based on the procedural model and incorporates key communication events that occur at each target: Whether the DEMPC informed the AVO and PLO of upcoming targets (e.g., restrictions, effective radius), whether the DEMPC was given information by the AVO or PLO, whether the PLO and AVO negotiated airspeed and altitude at the target, and whether the AVO was told by the PLO that the photograph taken at the target was acceptable (thus indicating to the AVO that the team is clear to move to the next waypoint).

Experimenters were also able to indicate if a particular communication event did not occur, if a packet of information was re-passed, if they were not sure a particular event occurred (in order to review the videotape and make confirmations that the

event in question did or did not occur), and make comments at each particular target. The experimenter logged events in real-time while remotely observing the team and listening to the audio. Each time an observation was logged it was associated with a time stamp. In addition, team process ratings described in the next section were entered using this software. Interfaces have been developed for the text communication system (see Figure 19) and the voice communication system (see Figure 20). Although the two look different, they are functionally identical.

Figure 19. Coordination and process loggers used in the text communication condition.

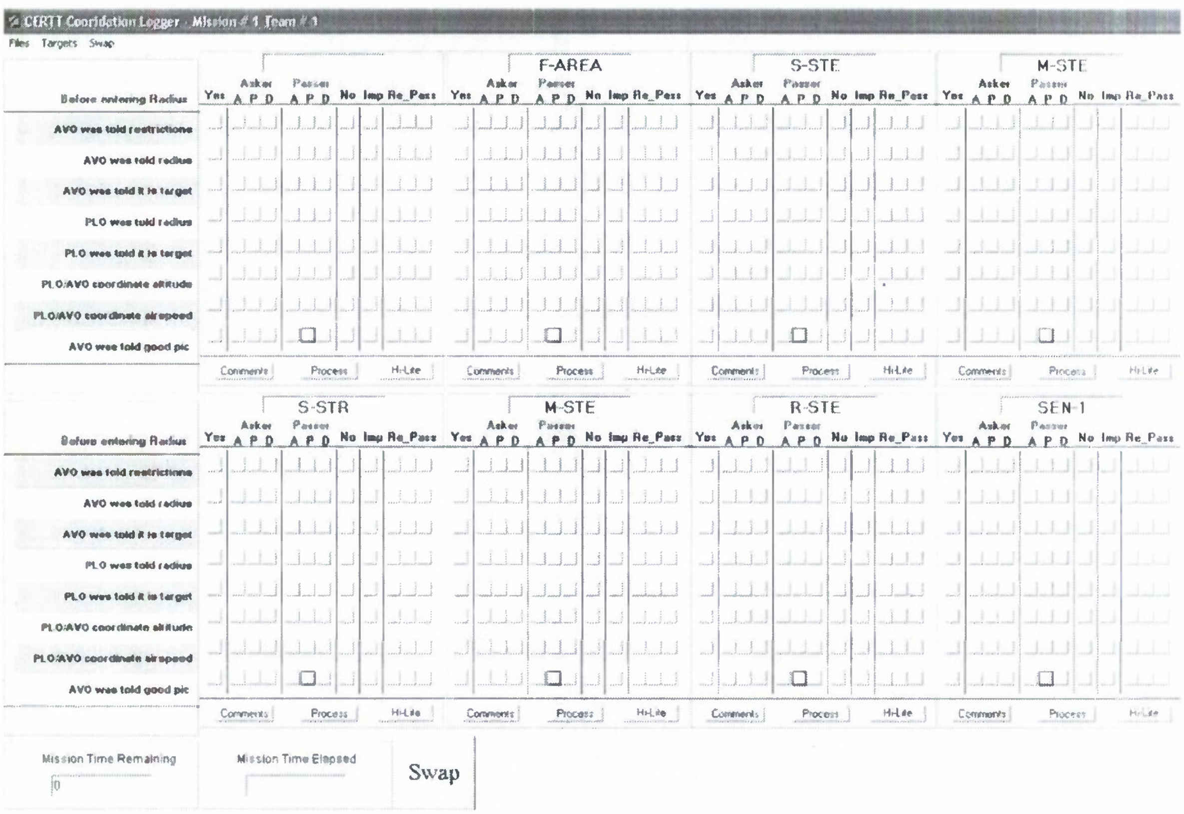

Figure 20. Coordination logger interface used in the voice communication condition.

<u>Team Process Rating</u>

Team process was scored by consensus between the two experimenters. For each target, the experimenters observed team behavior based on the key coordination events recorded on the coordination logger. The experimenters rated process on a scale ranging from 0 to 4 with 4 indicating "excellent" process and 0 indicating "poor" process. The rating was based on the timing of communications, number of repeated communications, situation awareness behaviors, and whether the team followed and included all elements of the procedural model for that particular target.

Process ratings reflect the experimenters' evaluation of team process behaviors, conceptualized as the level of coordination/communication, timeliness of interactions, team situation awareness, and overall impressions of the team acting as a well-integrated behavioral unit. DVD recordings and text communications for ten percent of all missions (n = 10 missions) were coded (using the coordination logger) independently by separate experimenters in order to assess inter-rater agreement.

To assess reliability among team process raters 10 missions composed of 70 targets were randomly selected to be independently rated by a second experimenter. ICC (Intraclass Correlation Coefficient) was calculated. The results of the of the analysis indicate that raters were in agreement (ICC = .71, $F(69, 69) = 3.419$, $p < .01$).

A 2 (communication condition) x 4 (mission) mixed ANOVA was conducted to determine if there were differences in process ratings as a function of communication mode or experience. There was a main effect of mission, $F(3, 51) = 8.72$, $p < 0.001$, where process ratings improved with experience. There was not a main effect of communication mode, $F(1, 17) = 2.37$, $p < 0.142$, indicating that the voice communication condition ($M = 2.39$) did not significantly differ from the text communication condition ($M = 1.88$), (see Figure 21).

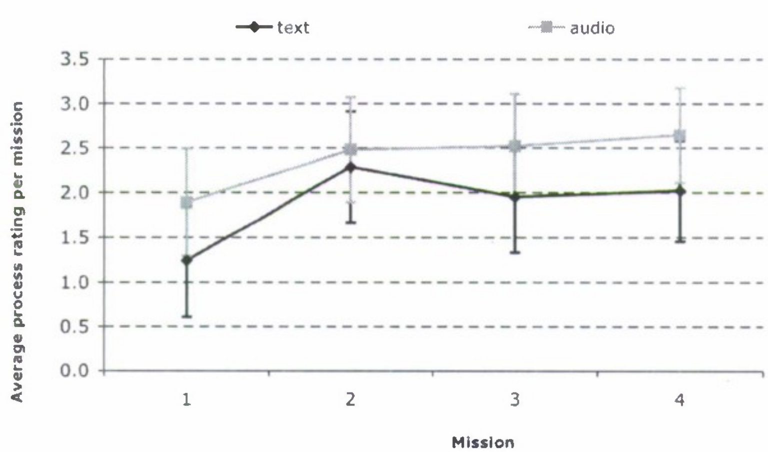

Figure 21. Average process ratings across missions and communication conditions. Error bars represent 95% confidence intervals.

To determine how the high workload mission affected process ratings across communication conditions, a 2 (communication condition) x 2 (mission) mixed

ANOVA was conducted. There was a significant main effect of workload on process ratings, where the high-load mission (M = 2.02) had significantly lower process ratings than from the low-load mission (M = 2.33), $F(1, 18) = 5.05$, $p = 0.037$. There was not a main effect of communication condition $F(1, 18) = 2.79$, $p = 0.112$, nor was there a significant communication condition x workload interaction, $F(1, 18) = 0.142$, $p = 0.711$.

Dynamical Systems Models of Team Coordination

The overall objective of this part of the work was to develop a dynamical systems model of team coordination with control parameters for determining possible differences in team coordination due to communication conditions. Sub-goals for achieving the overall objective included conceptualizing the fundamental nature of team coordination as a dynamical system, identifying a model (or set of models) that apply to this conceptualization, and evaluating the results of the experiment with reference to the model.

These analyses were still in progress at the time this report was written.

Bulleted Results Summary

This section contains a bulleted summary of all of the analyses presented above.

Performance:

- Team performance increased with experience.
- The main effect of communication mode (text, voice) did not significantly affect team performance ($p = 0.46$).
- Load affected team performance, where team performance decreased with increased load, as expected.
- Across the three roles, PLO was the only role to demonstrate an effect of communication mode on performance, with PLO participants in the voice communications condition performing better than PLO participants in the text-based communications condition.

Subjective Workload:

- The DEMPC perceives the greatest amount of mental demand
- The AVO experiences the greatest physical and temporal demands from the task.
- PLOs and AVOs from the text-communication condition experience greater temporal demand as workload increases
- PLOs and AVOs from the voice-communication condition experience the same and less temporal demand, respectively.
- The DEMPCs in each communication condition maintain stable levels of temporal demand as workload increases.

Communication Synchronicity

- The text-based communication condition functioned as an asynchronous communication platform across teammates and missions.
- The DEMPC reduced time lags across missions at a greater rate than the PLO or AVO

Taskwork Knowledge

- For *interpositional knowledge accuracy*, teams showed a significant increase in Taskwork Knowledge from Session 1 to Session 2.
- For *intrateam similarity,* teams showed a significant increase in Taskwork Knowledge from Session 1 to Session 2
- The increases in Taskwork Knowledge are attributable to increased communication and knowledge gathering about other team-members' roles.

Teamwork Knowledge

- For *overall accuracy*, text communication condition teams showed a significant decrease in teamwork knowledge from Session 1 to Session 2.
- For *interpositional knowledge accuracy*, text communication mode teams showed a significant decrease in teamwork knowledge from Session 1 to Session 2.
- Decreases in the above may be attributable to limitations in the amount of communication possible in the Text Chat environment as well as the fact that the AVO was not co-located.

Team Process

- Process ratings increased from mission 1 to mission 2, where it appears to have reached asymptote.

- There were no differences between communication conditions

Conclusions

The results from the first experiment demonstrated that text-based communications do not produce a reliable effect on team performance and team process when compared to voice-based communications. However, individual teammate performance analyses demonstrated that the PLO was negatively affected by the text-based communication system. Not surprisingly, the text-based communications is an asynchronous communication system. Consequently, we expect team coordination to change due to system asynchrony, and these analyses were being conducted at the time this report was written.

Synthetic Teammate Overview of Modeling Effort

The synthetic teammate is a functioning and cognitively plausible agent capable of interacting with humans to perform the UAV reconnaissance task. The synthetic teammate is being developed within the ACT-R cognitive architecture (Anderson & Lebiere, 1998; Anderson et al., 2004, Anderson, 2007), reflecting the focus on cognitive

plausibility. The constraints imposed by the architecture push system development in cognitively plausible directions which are more likely to lead to human-like behavior than purely algorithmic solutions which ignore such constraints

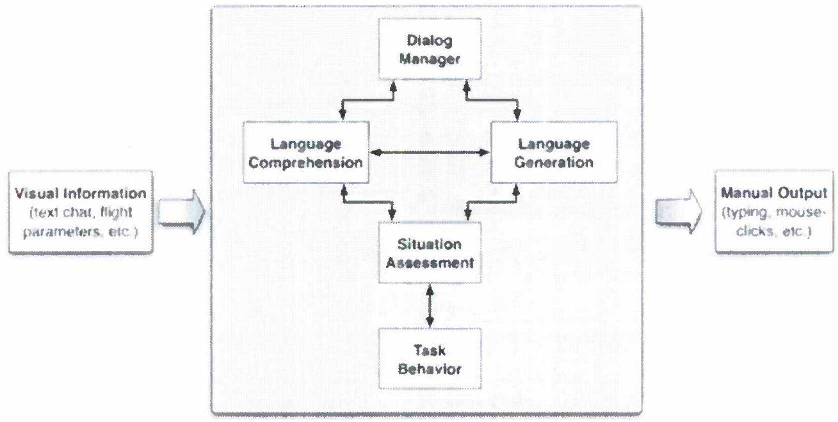

Figure 22. Synthetic teammate system overview.

The major linguistic components of the system include text-based language comprehension and generation components, which are under the control of a dialog manager (see Figure 22). The linguistic subsystem interacts with a situation assessment component that is a spatial /propositional representation of the current state of affairs as encoded from environment interactions (e.g., communications, flight controls, etc.). The situation assessment component functions to link linguistic representations from the language comprehension component to state representations from other components, and provides the interface for language comprehension and generation to the task behavior component.

The task behavior component implements the behavior of the system, controlling shifts of attention in the visual system and motor actions needed to perform the pilot's tasks. Input to the system is mediated by ACT-R's perceptual module and motor actions are mediated by ACT-R's motor module. The perceptual and motor modules are ACT-R's interfaces to the external environment. Each of the model components makes use of ACT-R's declarative and procedural memory systems. The following sections will provide more detail for each of the synthetic teammate's core components.

Language Comprehension Component

The language comprehension component is intended to be a domain general system capable of handling a wide range of English constructions (Ball, 2007a) based on an underlying linguistic theory of the grammatical encoding of referential and relational meaning (Ball, 2007b). Lexical items in the linguistic input activate constructions that drive processing.

The language comprehension component processes the input incrementally (one word at time), constructing a linguistic representation of the input based on the current word, constructions activated by the word, and the prior context. If necessary, the current input is accommodated by adjusting the current representation or coercing the current input

into that representation without backtracking. The mechanism of context accommodation is part and parcel of the basic left-to-right, incremental processing mechanism and is not viewed as a separate repair mechanism. The language processor is highly context sensitive and makes use of all available information—lexical, syntactic, semantic and pragmatic—in deciding how to process a given input. There is no autonomous syntactic component or syntactic processor, although grammatical information is very important for determining meaning.

The context sensitivity of the language processor makes possible a nearly deterministic processing mechanism. Contextual information is probabilistically summed via ACT-R's parallel spreading activation mechanism to yield the best alternative given the current input and context. This alternative is assumed to be correct and the processor proceeds deterministically and serially forward. Context accommodation provides a mechanism for dealing with the situation where the context and input leads to a choice that is locally preferred, but not globally preferred, adjusting the evolving representation without backtracking. The context sensitive, probabilistic, parallel, spreading activation substrate, combined with a mechanism of context accommodation makes a nearly deterministic, serial language processing system possible.

Language Generation & Dialog Manager Component

The language generation and dialog manager component was developed to capture the dynamic nature of human language production, following earlier approaches involving dynamic dialogue constraints (Ericcson, 2004), accommodation (Matessa, 2000), and adaptive content selection (Walker et al., 2004). The focus of the model is on selecting from a set of possible utterances, akin to overgeneration-and-ranking approaches (Varges, 2006).

The model uses optimality theory (Prince & Smolensky, 1993; 2004) to select an optimal utterance, given a set of utterances and a set of constraints on utterances. Constraints are simple, violable, conflicting, and motivated by cross-linguistic evidence. Constraints are arranged in a strict dominance hierarchy; the optimal utterance is the one that least violates the hierarchy.

Constraint ranking is expressed through ACT-R declarative memory activation: the most important constraint is most highly activated. Activation spreads from constraints to utterances to determine the utterance retrieved from memory; the most important constraint has the greatest effect on the retrieval. Factors from the situation component dynamically affect the constraint ranking, providing a principled variation in utterances over time.

Task Behavior Component

The task behavior component was developed to fly the UAV from waypoint to waypoint in a cognitively plausible manner. Flying to waypoints involves interacting with the UAV-STE to queue the correct waypoint and enter the correct course. The pilot must also set the UAV airspeed and altitude within restrictions provided by the sensor operator (PLO) and planning officer (DEMPC). The task model interacts with the UAV-STE using the same devices as humans–it uses the mouse pointer to interact with the UAV

flight controls in a point-and-click fashion, and uses the keyboard to send and receive messages to and from its teammates.

The task model was developed using a combination of hierarchical task analysis and NGOMSL notation (Kieras, 1988). The analysis identified the goals necessary for accomplishing flight from one waypoint to another, the sequence flexibility of the goals, and commonalities across all goals.

The task behavior goals associated with the task model include setting flight parameters (i.e., altitude, speed, and course), setting waypoints, monitoring alarms and warnings, and monitoring the UAV flight status (i.e., the distance from upcoming waypoint and the time to the next waypoint, etc.). Each of these goals was divided into three subgoals, checking current state information, obtaining desired state information, and changing the current state to the desired state. Each subgoal updated the appropriate information within the situation component.

The first component, checking, was modeled to obtain the current state information and determine if it differed from the desired state. When there was a discrepancy, the model performed the second component, obtaining, to get the desired state information from memory, the GUI, or one of its teammates. On obtaining the correct information, the model performed the third component, changing, to modify the task to a desired state. As a result of breaking each of the task goals into three components, there has been a substantial re-use of production rules within the task model.

For example, assume the task behavior component has received the next waypoint from the planning officer. This information is stored in the situation assessment component from the language comprehension component, and used to retrieve the goal from memory for checking waypoint information. To check the next waypoint value, the model attends and encodes the "queued waypoint" value on the GUI and determines if the queued waypoint needs to be adjusted. If the waypoint needs to be adjusted, then the task model spawns a goal to obtain the necessary information from memory, the GUI, or its current situation representation. Once the information is obtained, the task model attends the waypoint setting information and sets the desired waypoint using the appropriate mechanism.

Situation Assessment Component

The situation assessment component provides the interface between the linguistic components and the task behavior component. The situation assessment component is responsible for grounding the meaning of referring expressions and for representing the task environment. This component constitutes the primary meaning representation for the system. It is intended to have spatial and propositional properties. Within this context, we are evaluating a range of theories for use as the representational basis for constructing the situation assessment component, including Situation Models, Mental Models, Mental Spaces, Discourse Representation Theory, Discourse Space Theory and Conceptual Semantics. A key shortcoming of many of the identified theories is the lack of an embodied basis for representing meaning and the exclusive reliance on essentially propositional (and we think, linguistic) representations. However, we have not identified

any non-robotic, embodied approaches to meaning representation that provide the basis for a computational implementation. Using robots with sensors and a visual system, Mavridis & Roy (2006) are able to ground meaning in a situation model; and Scheutz, Eberhard & Andronache (2004) also use robots with sensors to ground meaning. The most recent ACT-R theory does not yet provide a full visual system capable of grounding meaning, but Douglass (2007) used the theory to develop a model of situated action. His work showed that situated actions based on active perception utilizing learned visual routines can be modeled using symbolic representations and rules in ACT-R. We plan on using this work to develop spatial representations for grounding information such as waypoint lists. We also find the idea of replacing the use of uppercase words corresponding to concepts (which are clearly linguistic) with iconic representations attractive.

Scaling up the ACT-R Cognitive Architecture

The ACT-R cognitive architecture was designed to support the development of small-scale cognitive models of specific laboratory phenomena. Since the advent of the first computational version of ACT-R, hundreds of small-scale models have been developed. The synthetic teammate project is one of a few attempts to develop a larger-scale model (or system of models) in ACT-R. This development is pushing the architecture in directions for which it was not originally designed. For example, the parallel spreading activation mechanism of the ACT-R architecture is computationally explosive on serial hardware. To support the computation of the activation of declarative memory chunks corresponding to thousands of lexical items, we have integrated the PostGreSQL relational database with ACT-R. The database provides a mechanism to externalize ACT-R's declarative memory and efficiently retrieve stored memories. Integration of the database also supports retrieval of lexical items based on the letters, bigrams and trigrams in the lexical item, instead of requiring a full-word match. This capability is needed for dealing with the variability in the input form of many lexical items in our text communications corpus and is also more cognitively plausible. Finally, the integration of a relational database allows us to easily build and maintain declarative knowledge acquired over many model runs.

Empirical Validation

An important goal of the project is to develop a synthetic teammate that is at once functional and cognitively plausible. In a system as complex as the synthetic teammate, empirical validation is a significant challenge. It is impractical to individually validate all system behaviors. Instead, a few key behaviors will be selected for scrutiny and validated against empirical data. At the highest level, we will determine whether or not teams with a synthetic AVO show evidence for the basic learning effect characteristic of all human teams in the UAV-STE. We also plan to compare the communicative behavior of the synthetic teammate in terms of the "push" and "pull" of information against data that has been collected for human teams. It should be noted that this empirical validation will occur within the context of a functioning synthetic teammate, an atypical empirical approach which will lend credibility to the model in the sense that the model must do much more than just show evidence for aligning with a specific data set – the model must

also function as a teammate with all the constraints on model development that that entails.

Furthermore, it is an empirical goal of the language comprehension component to be able to process linguistic input in real-time on Marr's algorithmic level (Marr, 1982) where parallel and serial processing mechanisms are relevant (Ball, 2008). This goal imposes serious constraints on possible processing mechanisms—for example, eliminating non-deterministic mechanisms that rely on algorithmic backtracking and cannot, in principle, operate in real-time since such mechanisms slow down with the length of the linguistic input.

Finally, not all components of the synthetic teammate are equally cognitively plausible. In the interest of building an end-to-end system, cognitive constraints on the development of the language generation and dialog manager components have been relaxed. On the other hand, the task behavior component, which takes advantage of the perceptual-motor modules of the ACT-R cognitive architecture, is closely tied to cognitive plausibility—down to the timing of keypresses and mouse movements.

Model Validation

The validation effort for the synthetic teammate has started, but is far from complete. To fully validate the model, it must be capable of completing five consecutive missions with human teammates, and is a focus of future research. The task behavior component has been preliminarily validated against human data

To fly the UAV from waypoint to waypoint in the CERTT task, a pilot must complete several goals, identified in the NGOMSL analysis. The key goals for piloting the UAV are checking and setting a queued waypoint, checking and setting a new waypoint, and checking and setting the course, altitude, and airspeed.

The three dependent variables compared between humans and the task behavior model were (1) the number of actions to complete a goal, (2) the time to complete a goal, and (3) the time between clicks when performing mouse clicks to complete a goal. Dependent variables one and two provide an accuracy estimate of the strategy implemented in the task behavior component for completing each of the goals, and the third dependent variable provides an accuracy estimate of low level motor times modeled within the ACT-R architecture.

Five humans participated in providing baseline data for each of the aforementioned goals. Human participants other than those used in the experiment were used because the CERTT AVO station does not collect data at the button or mouse-click level of analysis. Consequently, the five human participants set the airspeed, course, altitude, and new waypoint settings twenty times, each. Ten model runs were performed where each model run set the airspeed, course, altitude, and new waypoint settings twenty times, each. There was a low root mean squared deviation between model and human mean setting durations (i.e., 0.99, see Figure 23), mean number of actions (i.e., 1.84, see Figure 24), and the mean duration between clicks (i.e., 0.03, see Figure 25).

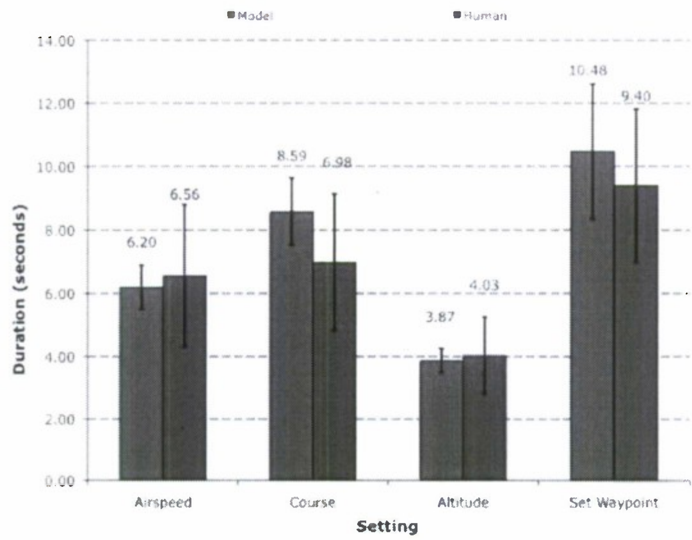

Figure 23. Human and model mean setting durations.

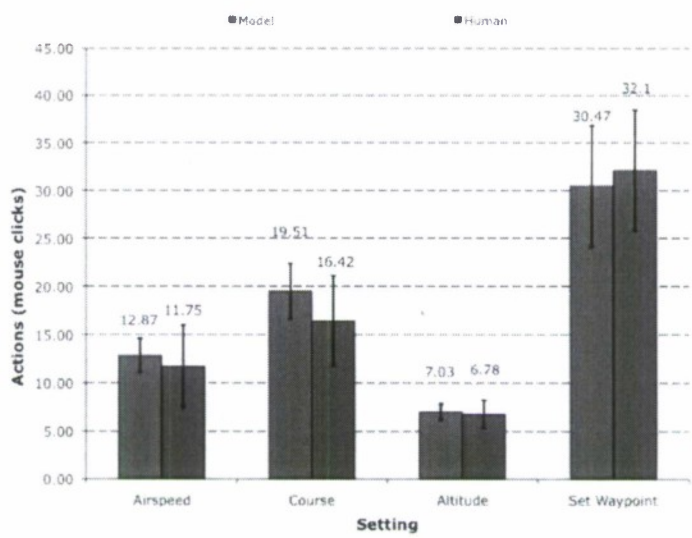

Figure 24. Human and model mean number of actions.

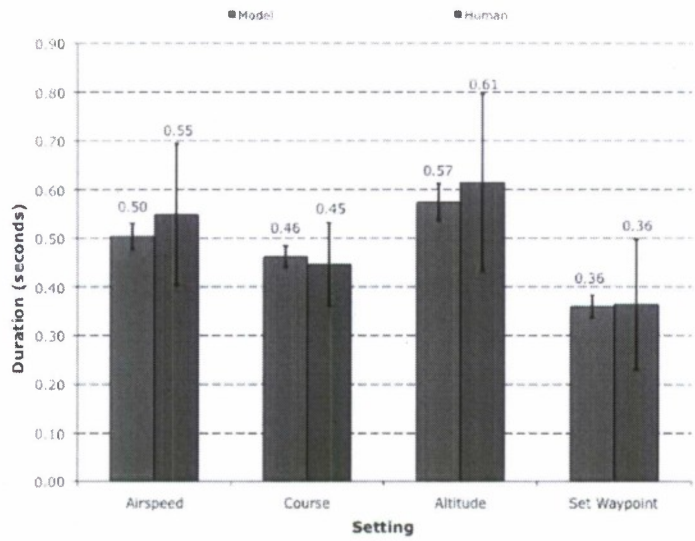

Figure 25. Human and model mean inter-click durations.

These results indicate that the task behavior model is an accurate representation of human behavior when completing these goals.

Conclusions

The Synthetic Teammate project is a challenging project reminiscent of earlier research in Artificial Intelligence and Cognitive Science that focused on solving AI Hard Problems using cognitively motivated computational techniques. The current goal is to have an initial end-to-end system in place by summer 2009. The initial system will be subjected to iterative refinement until a version that is capable of functioning as a teammate in the UAV-STE simulation is available. The research is guided by well established cognitive constraints on human language and task behavior and the system will ultimately be empirically validated against human performance data at the individual and team levels.

Contributions

Given that funding was discontinued after the second year, tasks associated with the third year of funding have been removed.

OBJECTIVE 1.0: Conduct Empirical Study of Cognitive Coordination to Guide Development of Synthetic Teammate

Task 1.1 Modify synthetic test bed to accommodate chat-only communications (1) **COMPLETED**

Task 1.2 Design Experiment 1 (chat vs. voice communications Agent) (1) **COMPLETED**

Task 1.3 Collect Experiment 1 data (1) **COMPLETED**

Task 1.4 Analyze and report Experiment 1 (2) **COMPLETED**

OBJECTIVE 2.0: Develop Synthetic Teammate

Task 2.1 Conduct task analysis of AVO performing reconnaissance task (1) **COMPLETED**

Task 2.2 Develop plan for staging Synthetic AVO development for mitigation of risk (1) **COMPLETED**

Task 2.3 Develop an interface between the CERTT simulation environment and ACT-R/Lisp (2) **COMPLETED**

Task 2.3.1 Visual input to Synthetic AVO (2) **COMPLETED**

Task 2.3.2 Data interface to support reimplementation of AVO GUI in ACT-R/Lisp environment (2) **COMPLETED**

Task 2.3.3 Motor output from Synthetic AVO (2) **COMPLETED**

Task 2.4 Develop Cognitive Model (reconnaissance task, cognitive control, reading, typing, comprehension of situation, cooperative dialog, representing other minds) (2-3) **ONGOING**

OBJECTIVE 3.0: Conduct an Empirical Study to Validate Synthetic Teammate and Test Coordination Training

Task 3.1 Incorporate Synthetic Teammate in synthetic test bed (2) **COMPLETED**

Publications & Presentations

Ball, J., Myers, C. W., Heiberg, A., Cooke, N. J., Matessa, M., & Freiman, M. (2009). The Synthetic Teammate Project. In the proceedings of the 18th Annual Conference on Behavior Representation in Modeling and Simulation. Sundance, UT.

Cooke, N. J. & Myers, C. W. (2008). An ACT-R Model of a Synthetic Teammate. Invited paper presented at the Developing and Understanding Computational Models of Macrocognition. Havre de Grace, MD.

Myers, C. W., Cooke, N. J., Ball, J. T., Heiberg, A., Gluck, K. A., & Robinson, F. E. (under review). Collaborating with Synthetic Teammates. In W. Bennett (ed.), Collaboration in Complex Task Environments

Myers, C. W., Gorman, J., Duran, J. L., & Cooke, N. J. (in preparation). Differences in Coordination Dynamics Between Synchronous and Asynchronous Communication Systems in a Team Task.

References

Alligood, K.T., Sauer, T.D., and Yorke, J.A. (1996). *Chaos: An introduction to dynamical systems.* New York: Springer-Verlag Inc.

Alonso, E. (2002). AI and Agents: State of the Art. *AI Magazine*, Vol 23, pp. 25-30.

analyzing student interactions in computer mediated communication environments. *Journal of Interactive Learning Research, 12*(1), 41-68. Prinz, J. (2002). *Furnishing the Mind.* Cambridge, MA: The MIT Press.

Anderson, J. R. & C. LeBiere (1998). *The Atomic Components of Thought.* Hillsdale, NJ: LEA.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. *Psychological Review 111*, 1036-1060.

Artman, H. (2000). Team situation assessment and information distribution. *Ergonomics, 43*, 1111-1128.

Badler, N., Phillips, C. & Webber, B. (1993). *Simulating Humans.* NY: Oxford University Press.

Ball, J. (2004). A Cognitively Plausible Model of Language Comprehension. In *Proceedings of the 13th Conference on Behavior Representation in Modeling and Simulation*, pp. 305-316. ISBN: 1-930638-35-3

Ball, J. (2006). Can NLP Systems be a Cognitive Black Box? In *Papers from the AAAI Spring Symposium, Technical Report SS-06-02*, pp. 1-6. Menlo Park, CA: AAAI Press

Ball, J., Gluck, K., Krusmark, M., and Rodgers, S. (2003). Comparing Three Variants of a Computational Process Model of Basic Aircraft Maneuvering. In *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*, pp. 87-98. ISBN 1-930638-30-2.

Barsolou, L. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences*, Vol 22, pp. 577-660.

Brannick, M. T., Prince, A., Prince, C., and Salas, E. (1995). The measurement of team process. *Human Factors, 37*, 641-651.

Busemeyer, J. R. (2002). "Dynamic Systems." *Encyclopedia of Cognitive Science.* L. Nadel (ed.). Macmillan Publishing.

Carpenter, P., Just, M., Keller, T., Eddy, W., Thulborn, K. (1999). Time Course of fMRI-Activation in Language and Spatial Networks during Sentence Comprehension. *Neuroimage*, Vol 10, pp. 216-224.

Cassell, J., Sullivan, J., Prevost, S. & Churchill, E. (2000). *Embodied Conversational Agents*. Cambridge, MA: The MIT Press.

Cassimatis, N. L. (2002). "Polyscheme: A Cognitive Architecture for Integrating Multiple Representations and Inference Schemes." Doctoral Dissertation, Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Cassimatis, N. L. (2005). "Integrating Cognitive Models Based on Different Computational Methods. *Proceedings of the Twenth-Seventh Annual Conference of the Cognitive Science Society.*

Cassimatis, N. L., Trafton, J. G., Bugajska, M. D. & Schultz, A. C. (2004). "Integrating cognition, perception and action through mental simulation in robots." *Journal of Robotics and Autonomous Systems.* Vol 49, pp. 13-23.

Chapman, R. J., Ryder, J., and Bell B. (2004). STRATA (Synthetic Teammates for Real-time Anywhere Training and Assessment): An integration of cognitive models and virtual environments for scenario based training. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting,* New Orleans, Louisiana. Santa Monica, CA: HFES

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press. communication medium on team interaction styles and outcomes. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(8-B), 4521-4718.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies, 41,* 801-849.

Cooke, N. C., & Gorman, J. C. (in press). Assessment of team cognition. In the 2nd edition of the *International Encyclopedia of Ergonomics and Human Factors.* Boca Raton, FL: CRC Press.

Cooke, N. J., Kiekel, P. A., Bell, B., & Salas, E. (2002). Addressing limitations of the measurement of team cognition. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 403-407.

Cooke, N. J., Kiekel, P. A., & Helm E. (2001). Comparing and validating measures of team knowledge. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.*

Cooke, N. J., Kiekel, P.A., & Helm E. (2001). Measuring team knowledge during skill acquisition of a complex task. *International Journal of Cognitive Ergonomics: Special Section on Knowledge Acquisition*, 5, 297-315.

Cooke, N. J., Rivera, K., Shope, S.M., & Caukwell, S. (1999). A synthetic task environment for team cognition research. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, 303-307.

Cooke, N.J., Salas, E., Kiekel, P.A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas and S. M. Fiore (Eds.), *Team Cognition: Understanding the Factors that Drive Process and Performance*, pp. 83-106, Washington, DC: American Psychological Association.

Cooke, N. J. & Shope, S. M. (1998). Facility for Cognitive Engineering Research on Team Tasks. Report for Grant No. F49620-97-1-0149.

Cooke, N.J. & Shope, S.M. (2002). Behind the scenes. *UAV Magazine*, 7, 6-8.

Cooke, N. J., & Shope, S. M. (2002). The CERTT-UAV Task: A Synthetic Task Environment to Facilitate Team Research. *Proceedings of the Advanced Simulation Technologies Conference: Military, Government, and Aerospace Simulation Symposium*, pp. 25-30. San Diego, CA: The Society for Modeling and Simulation International.

Cooke, N.J., & Shope, S. M. (2005). Synthetic Task Environments for Teams: CERTTS's UAV-STE Handbook on Human Factors and Ergonomics Methods, pp. 46-1-46- 6. Boca Raton, FL: CLC Press, LLC.

Cooke, N. J., Shope, S. M., & Kiekel, P. A. (2001). Shared-Knowledge and Team Performance: A Cognitive Engineering Approach to Measurement Technical Report for AFOSR Grant No. F49620-98-1-0287.

Cooke, N.J., & Shope, S. M., & Rivera, K. (2000). Control of an uninhabited air vehicle: A synthetic task environment for teams. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, 389.

Cooke, N. J., Stout, R., & Salas, E. (1997) Expanding the measurement of situation awareness through cognitive engineering methods. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, 215-219.

Cooke, N. J., Stout, R., and Salas, E. (2001). A knowledge elicitation approach to the measurement of team situation awareness. In M. McNeese, E. Salas, and M. R. Endsley (Eds.), *New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments* (pp. 114-139). Santa Monica, CA: Human Factors and Ergonomics Society.

Cooke, N. J., Stout, R., Rivera, K., & Salas, E. (1998). Exploring measures of team knowledge. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 215-219.

Cooper, R. (2002). *Modelling High-Level Cognitive Processes.* Mahwah, NJ: LEA.

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review, 80*, 97-125.

Entin, E. E. and Serfaty, D. (1999). Adaptive team coordination. *Human Factors, 41*, 312-325.

Ericsson, K. A. & Kintsch, W. (1995). "Long-term working memory". *Psychological Review*, Vol 102, pp. 211-245.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117-140.

Fodor, J. A. (1983). *Modularity of Mind.* Cambridge, MA: The MIT Press.

Gibbs, R. (2006). *Embodiment and Cognitive Science.* NY: Cambridge University Press.

Gibson, C. (2001). From knowledge accumulation to accommodation: Cycles of collective cognition in work groups. *Journal of Organizational Behavior, 22,* 121-134.

Gluck, K. & Pew, R. (eds.) (2005). *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation.* Mahway, NJ: LEA.

Gluck, K. A., Ball, J. T., Gunzelmann, G., Krusmark, M. A., Lyon, D. R., & Cooke, N. J. (2005). A prospective look at a synthetic teammate for UAV applications. *Proceedings of the American Institute of Aeronautics and Astronautics Infotech@Aerospace Conference.*

Gluck, K. A., Ball, J. T., & Krusmark, M. A. (in press). Cognitive control in a computational model of the Predator pilot. To appear in W. Gray (Ed.) *Integrated Models of Cognitive Systems.* New York, NY: Oxford University Press.

Gluck, K., Ball, J., Krusmark, M., Rodgers, S. and Purtee, M. (2003) A Computational Process Model of Basic Aircraft Maneuvering. In *Proceedings of the Fifth International Conference on Cognitive Modeling*, pp. 117-122. Edited by F. Detje, D. Doerner and H. Schaub. Universitaets-Verlag Bamberg. ISBN 3-933463-15-7.

Gorman, J. C., Cooke, N. J., and Kiekel, P. A. (2004). Dynamical perspectives on team cognition. *In Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, Santa Monica, CA: Human Factors and Ergonomics Society.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36, 180-193.

Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J. & Petajan, E. (2002). "Creating Interactive Virtual Humans: Some Assembly Required." *IEEE Intelligent Systems*, July/August 2002, pp. 54-63.

Gray, W. (ed.) (in press). *Integrated Models of Cognitive Systems.* NY: Oxford University Press.

Gugerty, L. DeBoom, D., Walker, R., & Burns, J. (1999). Developing a simulated uninhabited aerial vehicle (UAV) task based on cognitive task analysis: Task analysis results and preliminary simulator data. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*, 86-90.

Gunzelmann, G. & Anderson, J. R. (2004)."Strategic Differences in the Coordination of Different Views of Space." *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: LEA.

Gunzelmann, G., Anderson, J. R., and Douglass, S., (2004). "Orientation Tasks Involving Multiple Views of Space: Strategies and Performance." *Spatial Cognition and Computation*, Vol. 4, No. 3, pp. 209-256.

Hambley, L. A. (2006). Virtual team leadership: The effects of leadership style and communication medium on team interaction styles and outcomes. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 66(8-B), 4521-4718.

Harnad, S. (1990). "The Symbol Grounding Problem." *Physica D*, Vol 42, pp. 335-346.

Henderson, J. & Ferreira, F. (eds.) (2004). *The Interface of Language, Vision and Action: Eye Movements and the Visual World.* NY: Psychology Press.

Hill, R. (1999). "Modeling Perceptual Attention in Virtual Humans." *Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation*. Orlando, FL: SISO, pp. 563-573.

Hill, R., Chen, J., Gratch, G., Rosenbloom, P. & Tambe, M. (1997). "Intelligent Agents for the Synthetic Battlefield". In *Proceedings of AAAI-97/IAAI-97*, pp. 1006-1012.

Hinsz, V. B. (1995). Group and individual decision making for task performance goals: Processes in the establishment of goals in groups. *Journal of Applied Social Psychology, 25*, 353-370.

Hinsz, V. B. (1999). Group decision making with responses of a quantitative nature: The theory of social decision schemes for quantities. *Organizational Behavior and Human Decision Processes, 80*, 28-49.

Holland, J. (1998). *Emergence from chaos to order*. Oxford University Press.

Hollingshead, A. B. and Brandon, D. P. (2003). Potential benefits of communication in transactive memory systems. *Human Communication Research, 29*, 607-615.

Hutchins, E. (1991). The social organization of distributed cognition. In L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 283-307). Washington, DC: American Psychological Association.

Johnson-Laird, P., *Mental Models,* Harvard University Press, Cambridge, MA, 1983.

Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P. G., & Koss, F. (1999). Automated intelligent pilots for combat flight simulation. AI Magazine, *20*(1), 27-41.

Juarrero, A. (1999). *Dynamics in Action.* Cambridge, MA: The MIT Press.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA: The MIT Press.

Kieras, D. E., and Meyer, D. E. (1997). "An Overview of the EPIC Architecture for Cognition and Performance with Application to Human-Computer Interaction." *Human-Computer Interaction, 12,* pp. 391-438.

Kintsch, W. (1998). *Comprehension, a Paradigm for C ognition.* NY: Cambridge University Press.

Kosslyn, S. M., (1994). *Image and Brain: The Resolution of the Imagery Debate,* Cambridge, MA: The MIT Press.

Laird, J., Jones, R. & Nielsen, P. (1998). Lessons learned from TacAir-Soar in STOW-97. In *Proceedings of the Seventh Conference on Computer Generated Forces and Behavioral Representation.* Orlando, FL

Laird, J., Newell, A. & Rosenbloom, R. (1987). "An architecture for general intelligence". *Artificial Intelligence,* Vol 33, pp. 1-64.

Langley, P. (in press). Cognitive architectures and general intelligent systems. *AI Magazine.*

Lee, F. J., & Anderson, J. R., (2001). "Does learning a complex task have to be complex? A study in learning decomposition," *Cognitive Psychology,* Vol. 42, pp. 267-316.

Leontev, D.A. (1990). Deyatelnost i potrefnost (Activity and need). In D.B. Davydov & D.A. Leontev (Eds.), *Deyatelnostnyi Podhod v Psihologii: Problemy i Perspektivy (The Activity Approach in Psychology: Problems and Perspectives)* (pp. 96-108). Moscow, USSR: APN.

Levin, E. & Passonneau, R. (2006). A WOz variant with contrastive conditions. Paper presented at INTERSPEECH 2006.

Logie, R. H. (1994). *Visuospatial Working Memory.* Hove, U.K.: LEA.

Losada, M.& Heaphy, E. (2004). The role of positivity and connectivity in the performance of business teams: A nonlinear dynamics model. *American Behavioral Scientist, 47,* 740-765.

Lyon, D. R. (2004). "Measuring Visuospatial Working Memory Using Path Visualization." L3 Communications Link Division Technical Memorandum 23-1, Mesa, AZ.

Lyon, D. R., Gunzelmann, G., and Gluck, K. A., "Emulating a Visuospatial Memory Field Using ACT-R." *Sixth International Conference on Cognitive Modeling*, Lawrence Erlbaum Associates, Mahwah, NJ, 2004, pp. 368-369.

Magnusen, J., Tanenhaus, M., Aslin, R. & Dahan, D. (1999). "Spoken Word Recognition in the Visual World Paradigm Reflects the Structure of the Entire Lexicon. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society.*

Malone, T. & Crowston, K. (1994). The Interdisciplinary Study of Coordination. *ACM Computing Surveys*, 26, 87-119.

Marslen-Wilson, W., & L. Tyler (1987). "Against Modularity." In *Modularity in Knowledge Representation and Natural-Language Understanding.* Edited by J. Garfield. Cambridge, MA: The MIT Press.

Mathieu, J. E., Goodwin, G. F., Heffner, T. S., Salas, E., and Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*, 273-283.

Mohammed, S. and Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across discipline boundaries. *Journal of Organizational Behavior, 22*, 89-106.

Newell, A. (1990). *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

Paivio, A. (1986). *Mental Representation, a Dual Coding Approach.* NY: Oxford University Press.

Pashler, H. (1998). *Attention.* Philadelphia: Psychology Press.

Pena-Shaff, J., Martin, W., & Gay, G. (2001). An epistemological framework for analyzing student interactions in computer mediated communication environments. *Journal of Interactive Learning Research*, 12(1), 41-68.

Purtee, M. D., Krusmark, M.A., Gluck, K. A., Kotte, S. A. & Lefebvre, A. T. (2003). "Verbal protocol analysis for validation of UAV operator model." *Proceedings of the 25th Interservice/Industry Training, Simulation, and Education Conference.* Orlando, FL: National Defense Industrial Association, pp. 1741-1750.

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology.* New York: Oxford University Press.

Roy, D. Hsiao, K. & Mavridis, N. (2004). Mental Imagery for a Conversational Robot. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics.* Vol 34, pp. 1374-1383.

Salas, E. Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Salvucci, D. D., Boer, R. R., & Liu, A. (2001). "Toward an integrated model of driver behavior in a cognitive architecture." *Transportation Research Record*, No. 1779, pp. 9-16.

Scheutz, Matthias, Eberhard, Kathleen, and Andronache, Virgil (2004) "A Real-time Robotic Model of Human Reference Resolution using Visual Constraints". *Connection Science.* Vol.16, pp. 145-167.

Schoelles, M., and Gray, W. D., (2000). "Argus Prime: Modeling emergent microstrategies in a complex simulated task environment," *Proceedings of the 3rd International Conference on Cognitive Modeling*, edited by N. Taatgen and J. Aasman, Universal Press, Veenendaal, Netherlands, pp. 260-270.

Scolaro, J. & Santarelli, T. (2002). Cognitive modeling teamwork, taskwork, and instructional behavior in synthetic teammates. In Proceedings of the Eleventh Conference on Computer-Generated Forces and Behavior Representation, (pp. 315-322). Orlando, FL: UCF Institute for Simulation & Training.

Shepard, R. and Metzler, J., (1971). "Mental Rotation of Three Dimensional Objects." *Science*, Vol. 171, pp. 701-703.

Steiner, I.D. (1972). *Group process and productivity.* New York: Academic Press.

Sun, R. (2001). Hybrid systems and connectionist implementationalism. *Encyclopedia of Cognitive Science.* MacMillan Publishing Company.

Sun, R. (2005). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In Ron Sun (ed.), *Cognition and Multi-Agent Interaction.* NY: Cambridge University Press

Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S, Rickel, J & Traum, D. (2006). Toward Virual Humans. *AI Magazine*.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, Vol 29, 557-580

Trueswell, J. & Tanenhaus, M (eds.) (2004). *World-situated language use: Psycholinguistic, linguistic, and computational perspectives on bridging the product and action traditions.* Cambridge, MA: MIT Press.

Tushman, M. L. (1979). Work characteristics and subunit communication structure: A contingency analysis. *Administrative Science Quarterly, 24,* 82-97.

Wermter, S. & Sun, R. (eds.) (2000). *Hybrid Neural Systems.* Heidelberg: Springer-Verlag.

Zachary, W., & Ross, L. (1991). Enhancing Human-computer interaction through use of embedded COGNET models. In Proceedings of Human Factors Society 35th Annual Meeting , (pp. 425-429). Santa Monica, CA: Human Factors and Ergonomics Society.

Zachary, W., Santarelli, T., Ryder, J., Stokes, J., & Scolaro, D. (2001). Developing a multi-tasking cognitive agent using the COGNET/iGEN integrative architecture. In Proceedings of 10th Conference on Computer Generated Forces and Behavioral Representation, (pp. 79-90). Norfolk, VA: Simulation Interoperability Standards Organization (SISO).

Zwaan, R. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In: B.H. Ross (Ed.), *The Psychology of Learning and Motivation,* Vol. 44 35-62. New York: Academic Press.

Zwann, R., and Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin,* Vol. 123, 162-185.

Appendices

Appendix A

Components of Individual and Team Performance Scores

Subscore	Subscore Numerator	Subscore Denominator	Transformation	Weight	Relati Weig
AVO					
Alarm Penalty	AVO Alarm Duration	missionTotalSecs	subscore^.5	126.69	4
Warning Penalty	AVO Warning Duration	missionTotalSecs	subscore^.5	25.14	1
Course Dev Penalty	From Flgt_Sum.rds, Sum of all "SumOfDev"	totalRouteLength	-	287.06	4
AVO Rte Seq Penalty	Planned WPs not Visited** + Visted WPs not Planned - WPs can't make*	total wps planned - WPs can't make*	-	262.94	3
PLO					
Alarm Penalty	PLO Alarm Duration	missionTotalSecs	subscore^.5	567.70	3
Warning Penalty	PLO Warning Duration	missionTotalSecs	subscore^.5	121.96	1
Duplicate Good Photos Penalty	totalGood - totalGoodUnique	film	-	1730.26	4
Missed or Slow Photo Penalty	totalGoodUnique	missionTotalSecs/60	1-subscore	39.02	2
Bad Photo Penalty	Bad Photos	Film	-	178.34	3
DEMPC					
Alarm Penalty	DEMPC Alarm Duration	missionTotalSecs	subscore^.5	265.93	2
Warning Penalty	DEMPC Warning Duration	missionTotalSecs	subscore^.5	30.93	1
Missed CWPs Not Planned Penalty	Critical WPs not planned	unique total wps planned	-	1200.6	4
Alarm WPs Penalty	Hazard/Lost WPs Planned	unique total wps planned	-	692.47	3
Rte Seq Plan Penalty	Rte Seq Plan Violation	total wps planned	-	1177.53	4
TEAM					
Alarm Penalty	TEAM Alarm Duration	missionTotalSecs	subscore^.5	393.22	2
Warning Penalty	TEAM Warning Duration	missionTotalSecs	subscore^.5	112.02	1
Missed or Slow Crit WPs Penalty	critical_reached	missionTotalSecs/60	1-subscore	318.63	3
Missed or Slow Photos Penalty	totalGoodUnique	missionTotalSecs/60	1-subscore	314.96	4

*WPs can't make = total wps planned - the number in the DEMPC route that signifies the last waypoint hit by AVO and planned by DEMPC

** Planned WPs not visited is not the same number as noted by the rapid file. It is the number of planned WPs not visited out of the unique WPs planned

Appendix B

Taskwork Ratings Task

Instructions: In this experiment you will be presented with pairs of items that are relevant to the team task that you have just completed. We would like you to rate each pair according to the degree of overall relatedness of the items in that pair. Two items can be related in a number of different ways. For example, you might base your rating on geographic proximity, similarity in outcomes, or similarity in causes. However, please do not dwell on specific dimensions like these. Instead, make your ratings based on your first general impression of relatedness.

Concept List (Presented in pairs):

Airspeed

Altitude

Effective Radius

Focus

Fuel

Mission Time

Photos

ROZ entry

Shutter speed

Target

Zoom

Appendix C

Teamwork Knowledge Questionnaire

Instructions: You will be reading a mission scenario in which your team will need to achieve some goal. As you go through the scenario in your mind, think about what communications are absolutely necessary among all of the team members in order to achieve the stated goal. For example, does the AVO ever have to call the DEMPC about something? Using checkmarks, indicate on the attached scoring sheet which communications are **absolutely necessary** for your team to achieve the goal.

Scenario: Intelligence calls in a new **priority target** to which you must proceed immediately. There are **speed and altitude restrictions** at the target. You must successfully **photograph the target** in order to move on to the next target. At a minimum, what communications are absolutely necessary in order to accomplish this goal and **be ready to move on to the next target?** (check those that apply)

__________AVO communicates altitude to PLO

__________AVO communicates speed to PLO

__________AVO communicates course heading to PLO

__________AVO communicates altitude to DEMPC

__________AVO communicates speed to DEMPC

__________AVO communicates course heading to DEMPC

__________PLO communicates camera settings to AVO

__________PLO communicates photo results to AVO

__________PLO communicates camera settings to DEMPC

__________PLO communicates photo results to DEMPC

__________DEMPC communicates target name to AVO

__________DEMPC communicates flight restrictions to AVO

__________DEMPC communicates target type (e.g., nuclear plant) to AVO

__________DEMPC communicates target name to PLO

__________DEMPC communicates flight restrictions to PLO

__________DEMPC communicates target type (e.g., nuclear plant) to PLO